POLITECNICO DI TORINO

Department of Electronics and Telecommunications

Master's Degree in Communications and Computer Networks Engineering

Study of 5G New Radio and performance analysis through numerical simulation of the Physical Downlink Control Channel

Supervisors

Prof. Roberto Garello

Candidate

Livia Francesca Rosa Marzo

April 2019

Thesis Project for the Master of Science in Communications and Computer Networks engineering

Study of 5G New Radio and performance analysis through numerical simulation of the Physical Downlink Control Channel

Livia Francesca Rosa Marzo

Prof. Roberto Garello

Department of Electronics and Telecommunications
POLITECNICO DI TORINO
Turin, Italy 2019

Summary

The purpose of this document is to provide a description of the methods adopted for the realization of the master thesis "Study of 5G New Radio and performance analysis through numerical simulation of the Physical Downlink Control Channel".

In the first chapter, it is set out an overview on the evolution of wireless communication technologies, with mention of the most important standards such as GSM, UMTS, LTE and New Radio.

The second chapter outlines the physical aspects involved in the structure of the 5th generation networks, from the numerology to the resource allocation, modulation techniques and transmission schemes.

The third chapter is focused on the Physical Downlink Control Channel of 5G NR describing its main features and structure. A specific section highlights the block chain of the control channel and in the following sections, two of the most innovative aspects are described: Polar Coding and Beamforming.

In the fourth chapter, a description of the MATLAB software implemented for the simulation of the PDCCH.

The fifth chapter reports the most significant results observed through the simulation sessions with considerations on the observed performances.

The final chapter is a brief discussion about the objectives set for the thesis project and the results actually obtained.

Contents

List of figures

1. Introduction

5G is the acronym commonly used to refer to the fifth generation of mobile communication that will succeed the ongoing fourth generation standardized through the LTE technology.

1.1 Brief history of wireless technology

Wireless communication systems are a constantly evolving process initiated almost 40 years ago and that includes more and more functionalities and application fields at each step [1].

Indeed, the mobile communication technology of 1st generation was introduced in the 1980s achieving very limited performances that simply included voice services. The main standards for 1G were the Nordic Mobile Telephone (NMT) used North-Europe, East-Europe and Russia, the Advanced Mobile Phone System (AMPS) used in North America and Australia, Total Access Communication System (TACS) in the United Kingdom and RTMI (Radio Telefono Mobile Integrato) in Italy.

Although the connection between radio towers and mobile equipment was implemented in digital domain, the rest of the signal transmission was led in the analogue domain so that the voice was only modulated to higher frequencies from 150 MHz on.

However 1G has been a very great novelty since it made mobile telephony accessible to ordinary people for the first time.

The second generation of cellular networks were introduced in 1990s when the GSM standard (Global System for Mobile communication) was deployed in Finland and rapidly spread out in many other countries even outside Europe.

Some other standards were released in different nations such as D-AMPS (Digital AMPS) and PDC (Personal Digital Cellular) in Japan but the dominant technology was GSM.

The substantial innovation of 2G was in the purely digital signaling along the entire voice transmission implying the digital encryption of the voice. This allowed to improve the quality of the voice and capacity of the network, furthermore 2G provided a supplementary data service for sending text messages which is commonly referred to as SMS.

The third generation of mobile communication developed starting from the 2000's as the UMTS (Universal Mobile Telecommunications Service) was standardized by 3GPP and spread out in Europe, Japan and China.

The main features deployed through 3G were: wireless voice telephony, mobile Internet access, fast and fixed wireless internet access, video calls and mobile TV. Throughout the years, 3GPP has developed several revisions of the UMTS technology, and introduced HSPA (High Speed Packet Access) protocols which enabled faster data connectivity both in downlink (with HSDPA) and in uplink (with HSUPA) through the introduction of a shared channel with higher order modulations support and fast physical layer adaptation to channel conditions.

The fourth generation of wireless communication was presented for the first time in 2009 as the Long Term Evolution (LTE) standards deployed in Sweden and even in some other countries of the world.

Starting from the requirements specified by ITU-R, LTE improved the performances experienced in the 3rd generation technology and succeeded in achieving higher efficiency.

New technologies and tools were exploited in 4G as the IP telephony where the communication is routed along datagram circuits ruled by Internet Protocols. Another cornerstone of 4G is the OFDM-based transmission enabling wider transmission bandwidths as well as the MIMO technology allowing additional improvements in the peak bit rate.

Furthermore, LTE supported both TDD and FDD transmission schemes in the same radio-access interface.

The idea of the fifth generation of the mobile technology has arisen in 2012 when the Working Party 5D (WP5D), within the ITU-R, published the ITU-R Recommendation M.2083 [2]. It was a recommendation introducing the IMT-2020, the International Mobile Telecommunications systems expected to be released in 2020.

The ITU-R Recommendation M.2083 [2] described the main features of the next generation systems at a very high-level.

Since then, the 3GPP started the process of the standardization of 5G and, as a result of this initial phase, it was launched the Release 15 in December 2017 that is considered the ‘’Early Drop’’.

This initial version of specifications described the application of 5G technology only in the field of the Enhanced Mobile BroadBand (eMBB) and it still used the LTE core network as control plane, so that 5G core network was a non-standalone radio technology.

After this first attempt, Release 15 has been revised and the URLLC (Ultra Reliable Low Latency Communications) was pointed out as another application of 5G and the stand-alone mode was presented.

A key point in the standardization process of 5G is the definition of a new radio-access technology that is called New Radio and that represents itself the fifth generation of the cellular networks.

1.2 5G enhancements

The main improvements introduced by 5G NR with respect to the LTE are [3]:

- *higher-frequency bandwidth*; in NR the licensed spectrum is from 1 GHz to 52.6 GHz, these are the so called mm-wave frequencies that allow very wide transmission bandwidths, increasing the capacity and data rates. The drawback is the increase of the attenuation implying the reduction of the coverage area that is partly overcome through the beam-centric design in NR. However, 5G will still exploit lower frequencies for some fundamental operations for the management of the network.
- *ultra-lean design*; related to the always-on signals exploited for the channel estimation. Depending on the average load this kind of signals may become the dominant traffic affecting the achievable network energy performance and the achievable data rates due to interference to other cells. Hence, the ultra-lean design reduces the always-on signals.
- *low latency transmissions*; achieved through several adjustments which have reduced transmission delays both in the uplink and downlink directions. For instance, in NR there is no time-domain interleaving along ODFM symbols and given that the reference signals are always at the beginning of the transmission, the receiver can process data immediately once they are received.
- *beam-centric design* exploiting beamforming and MIMO technologies for both data transmission and control-plane procedures; at higher-frequencies MIMO processing is exploited for a larger coverage, at low-frequencies for spatial-multiplexing.

1.3 Use cases

The macro application fields thought for the fifth generation of cellular networks are stated in the recommendation ITU-R M.2083 [2] and are essentially three: Enhanced Mobile Broadband, Ultra-reliable and low-latency communications, and Massive machine type communications.

Enhanced Mobile Broadband (eMBB)

it is the primary usage of 5G thought as the evolution of the 4G LTE mobile broadband services providing faster connections, higher throughput and higher capacity. This class is strictly related to the constant increasing of mobile data-traffic generated by more and more data-consuming devices coupled with multimedia applications. Hence, eMBB includes several usages finalized to a better communication experience from the human point of view improving hotspots for higher user density and extending the coverage for higher mobility scenarios.

Ultra-reliable and low-latency communications (URLLC)

the focus of this class is not only human-centric communication but there is a particular attention to machine connectivity exploiting wireless networks of the 5th generation, commonly referred to as IoT (Internet of Things). The requirements for this usage are mainly low latency, reliability, low or medium data rates, high speed mobility and high availability.

Some applications of the IoT are vehicle-to-vehicle communication for safety, wireless control of industrial equipment, remote medical surgery.

As regards human-centric URLLC, the most relevant are 3D gaming and “tactile internet” requiring also high data rates.

Massive machine type communications (mMTC)

this usage scenario is uniquely referred to machine-to-machine connectivity developing the IoT technologies. Everyday objects from different geographic regions will communicate each other through wireless networks. Therefore, it involves large number of transmissions of small data volumes not particularly sensitive to delays.

One of the fundamental requirements for mMTC devices is to keep costs low since it is expected a very large number of machines used in the environment and furthermore they will need very long battery life time.

Possible 5G applications

Going more into detail, some of the specific applications expected for the 5G networks are [3]:

- Autonomous vehicle control; basically designed to provide better traffic safety, to avoid accidents, for stress reduction and let drivers concentrate on other productive activities
- Emergency communication; thought to guarantee wireless communication in emergency situations when some part of the network may fail due to natural disasters
- Factory cell automation; to support life-critical applications automation
- High speed trains; providing high data rates even in the high mobility scenario as it is along rail
- Smart city; to connect humans to the surrounding environment providing personalized and location-aware services.
- Virtual and Augmented reality; to let people enjoy experiences even from remote locations such as gaming, meeting, playing music and conferences.

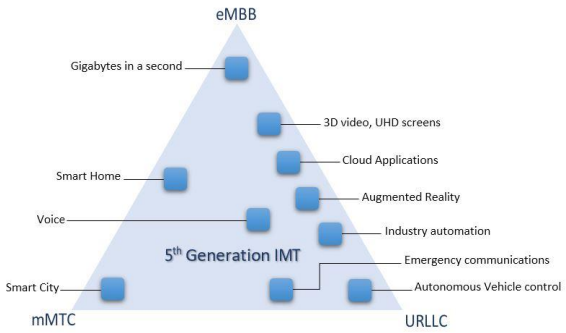


Figure 1.1 *5G use cases*

1.4 Capabilities

The recommendation ITU-R M.2083 [2] highlights 13 capabilities expected for the IMT-2020, eight of them are considered as key capabilities and are listed below.

Peak data rate

According to the ITU-R definition, peak data rate is the maximum achievable data rate under ideal conditions, hence without taking into accounts impairments due to real propagation channel. It depends on the available bandwidth and on the peak spectral efficiency. To achieve very high peak data rates it is necessary to have large bandwidths, but they're quite challenging to obtain at

frequencies lower than 6GHz, therefore the highest data rates will be mainly reached in indoor and hotspot environments using millimeter waves frequencies.

User experienced data rate

It quantifies the experienced data rate over a large cover area by the most part of the users. It is function of the available spectrum and of the type of the system. For 5G the target user experienced data rate is of 1 Gbit/s in both indoor and hotspot environments.

Spectrum efficiency

It measures the average data throughput per Hz of spectrum and per "cell" and is exploited for dimensioning networks. The aim of 5G is to reach a spectrum efficiency three times larger than the one experienced in 4G.

Area traffic capacity

It depends on the spectrum efficiency, the bandwidth and on the traffic density in a specific scenario.

Area traffic capacity= spectrum efficiency * BW * TRP density

Network energy efficiency

It is a capability to control the energy consumption and keep it as low as possible.

The challenge for 5G will be to reduce energy consumed per bit without impacting on transmission performances.

Latency

It measures the occurred time from the time when the source sends a packet to the time when it gets to the destination.

Mobility

Mobility is a capability that gained relevance when URLLC systems were conceived. It is a crucial parameter to be taken into account for granting the communication on high speed trains as well as for critical vehicle communication.

Connection density

It quantifies the total number of connected and/or accessible devices per unit area. Connection density is a particular issue for the mMTC systems where a lot of communicating machines are present in the environment.

Furthermore to these key capabilities, there are other five parameters highlighted in the ITU-R recommendation for IMT-2020 [2]:

Spectrum and Bandwidth flexibility

It is the capability of the system to adapt to different usages and different frequencies.

Reliability

It describes the ability to guarantee a high level of availability for a given service

Resilience

It is the aptitude of the network to keep on operating correctly in case of natural disasters or damages.

Security and privacy

It's a feature that has the aim to protect through encryption user data and signaling.

Operational lifetime

It is the amount of operation time per stored energy capacity.

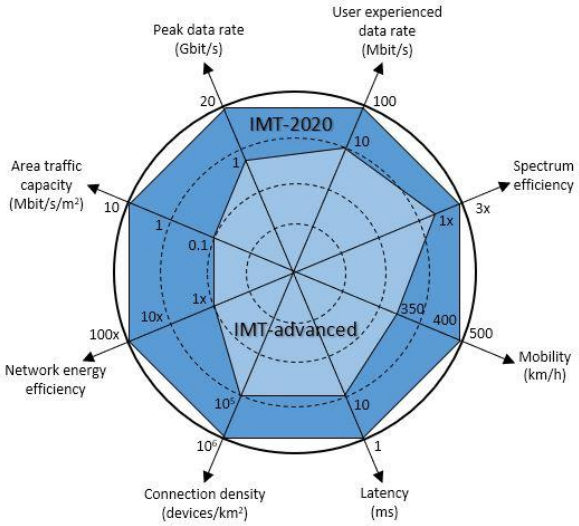


Figure 1.2 *5G key capabilities*

All of the capabilities mentioned before become more or less important depending on the particular service in which they are measured and it is difficult to optimize them simultaneously, indeed some of them are mutually exclusive.

2. Overview of the Physical Layer in 5G

The physical layer of a mobile network is the lowest level in the OSI structure, where raw bits are converted into a physical signal and subsequently transmitted along the wireless channel.

The 5G NR introduces several enhancements on the physical layer that enable the applicability to several fields with different requirements (low latency, ultra-reliability, high data-rates, low error rates,…).

As it is highlighted in [4], the main features of the physical layer are:

- Error detection and signalling to upper layers
- FEC encoding/decoding
- Hybrid H-ARQ soft combining
- Rate matching to the available resources on the channels
- Allocation of the transport channel onto physical channels
- Power measurements of the physical channels
- Modulation/demodulation techniques
- Time/frequency synchronization
- Implementation of MIMO (Multiple Input Multiple Output) transmissions
- RF management

The features mentioned above are designed in such a way to maintain a certain flexibility and to support forward compatibility and ultra-lean model.

2.1 5G spectrum

The spectrum of 5G technology includes new bands but also frequencies that were allocated for mobile networks of previous generations such as 3G and LTE.

It is possible to classify frequency bands of NR into three possible groups: low, medium and high frequencies [3].

Low-frequencies bands are those ones up to 2 GHz, they were previously exploited for the LTE channels. This type of frequencies provides very large coverage areas both in outdoor and indoor scenarios since the attenuation increases as the frequency gets higher.

Some usage cases may be traditional applications but also innovative tasks such as IoT applications as well as control systems for industries and business.

Bands at lower frequencies are not particularly broad, indeed they are typically 20 MHz wide. The most relevant bandwidths are the 600 MHz and the 700 MHz.

Medium-frequency bands are between 3 and 6 GHz, they are suitable for wide coverage, high data-rates and higher capacity requirements.

This range of frequencies can support larger bandwidths until 100 MHz.

In general, the most coveted bands are between 3.3 and 4.2 GHz, then they slightly change depending on the country, as for example in Europe the range is shrunk in 3.4-3.8 GHz, instead in China larger bands are coveted, hence in 4.8-5.0 GHz.

Part of medium-frequency bands are LTE frequencies that now have been allocated for NR.

High-frequency bands are over 24 GHz and, due to their propagation properties, they cannot provide wide coverage, hence their typical usage cases are hotspots deployment. Some considerable advantages are related to throughput performances, indeed high frequencies can achieve significant capacity and high data-rates. High-frequency bands are the very new frequencies, in the sense that it is in NR that they are exploited for the first time for IMT purposes.

Bandwidths up to 400 MHz are feasible and the bands of major interest are in 24.25-29.5 GHz.

Frequency-bands may also be classified according to their capabilities and their suitability to requirements [5]. As result, two ranges have been identified:

- *Frequency Range 1* (FR 1), including all the frequency bands lower than 6 GHz
- *Frequency Range 2* (FR 2), including all the new frequencies ranging from 24.25 and 52.6 GHz.

Bands in FR2 are the so-called millimetre wave (*mm-Wave*) frequencies and as mentioned before, they can reach a rate up to 40 Gbit/s.

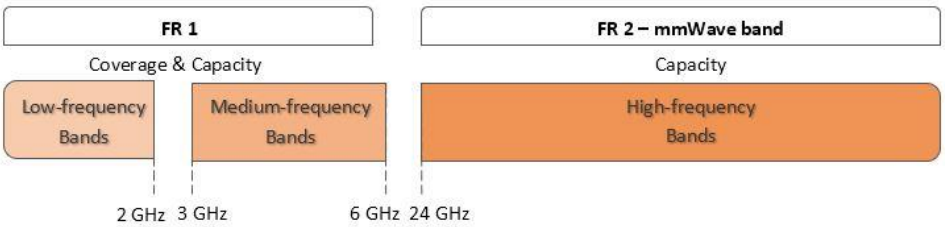


Figure 2.1 5G NR frequency Bands

2.2 Numerology

The numerology adopted in 5G is one of the most revolutionary aspects in the design of the fifth generation of mobile communications. The 3GPP has agreed on providing a scalable numerology in order to guarantee the adaptability of 5G to different scenarios with various requirements, sometimes even conflicting [6].

By means of scalable numerology, it is commonly denoted the spacing Δf between consecutive subcarriers depending on the parameter μ according to

$$\Delta f = 2^{\mu} \cdot 15\ kHz$$

where $\mu = 0,1,..,5$.

It can be noticed that the minimum spacing is $\Delta f = 15\ kHz$ that corresponds to the one adopted for LTE, hence in NR is still employed to maintain the backward compatibility. Larger values of spacing are obtained by multiplying by a factor of two the baseline subcarrier spacing at $15\ kHz$.

Another parameter strictly related to the subcarrier spacing is the cyclic prefix length, these two parameters impact on the performances over the channel.

Indeed, larger subcarrier spacing is suitable to reduce frequency errors and phase noise, but on the other side it impacts on the overhead due to cyclic prefix and for this reason it is better to reduce its duration. Moreover, the phase noise and the sensitivity to Doppler spread/shift are inversely proportional to the subcarrier spacing.

Hence subcarrier spacing has to be chosen as a trade-off between overhead due to cyclic prefix and the Doppler effect.

As an example, it can be considered a cell serving at mm-wave frequencies. In this scenario phase noise is not negligible and higher subcarrier spacings are needed. On the other hand, at mm-wave bands the cells have a small size and through the beamforming the delay spread is kept low. For these reasons, at higher frequencies, it is better to have larger spacings and shorter cyclic prefixes.

From a more practice point of view, it has been found that higher subcarrier spacings are suitable for URLLC applications, for small coverage areas and mm-wave bands. Smaller spacings support wider coverage areas, lower bands, narrowband devices and eMBMSs (evolved Multimedia Broadcast Multicast services).

Table 2-1 highlights the possible subcarrier spacing and the corresponding cyclic prefix length.

μ	Subcarrier Spacing (KHz)	Useful Symbol Time T_u (μs)	Cyclic Prefix $T_{CP}(\mu s)$
0	15	66.7	4.7
1	30	33.3	2.3
2	60	16.7	1.2
3	120	8.33	0.59
4	240	4.17	0.29

Table 2-1 *Possible numerologies in 5G NR*

Set of parameters (subcarrier spacing, cyclic prefix) with $\mu = 0,1,2$ are used for services working at bands below 6 GHz, the others are instead employed for mm-wave frequencies.

A further version for the cyclic prefix is the extended cyclic prefix that has a longer duration and is exploited when the delay spread is critical even if it impacts on the resulting overhead. The extended cyclic prefix can be used only when $\Delta f = 60\ KHz$, while at other spacings only normal cyclic prefix is deployed.

A relevant aspect is related to the scaling factor. In fact, multiplying by 2^{μ} ensures the alignment of different numerologies in the time domain, so that the duration of two OFDM symbols corresponds to one OFDM symbol at smaller spacing.

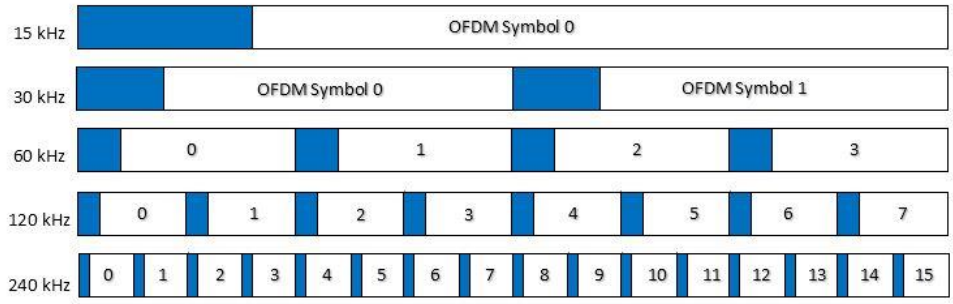


Figure 2.2 *NR Numerology*

2.3 Structure in time-domain

In NR, physical signals are transmitted over the time domain according to fixed size *frames* and *subframes* [7]. Each frame is 10 ms long, then it is furthermore divided into ten fixed size subframes of a length equal to 1 ms each.

Frames and subframes structure is equivalent to the one adopted for LTE so to guarantee the backward compatibility.

The flexibility of NR physical resources is embedded in the structure of subframes. Indeed, each subframe contain a variable number of slots depending on the parameter μ that determines the subcarrier spacing. Each slot consists of a fixed number of OFDM symbols which is equal to 14.

In the case of a spacing $\Delta f = 15kHz$, a subframe contains one slot, hence the duration of the slot itself fills the entire length of the subframe being 1 ms.

Increasing the spacing between subcarriers, the number of slots in a subframe increases and since the number of OFDM symbols within a slot is fixed, the result is that the duration of the slot itself decreases as well as the length of OFDM symbols.

A system subframe number (SFN) is exploited in order to identify each subframe especially in case of transmissions that cover more than one subframe.

A remarkable aspect has to be pointed out as concerns the synchronization in time domain. Indeed, since slots have variable size, they cannot be used as a time reference. Conversely, frame and subframes are not affected by the scaling factor, therefore they are deployed for coordination in the time domain among physical channels which may lead different numerologies.

It can be easily noticed that larger spacings are a suitable solution for low latency transmissions as URLLC. Indeed, increasing the spacing and consequently the number of slots per subframe, the length of slots is reduced.

Another possible solution to support low latency requirements consists in making the transmission independent from the slot period, in such a way that a packet delivery occupies only the number of OFDM symbols needed for the transmission of the payload. The latter scheme is defined as *mini-slot* and it includes two or more symbols, where the first one carries control signals.

Furthermore, mini-slots are beneficial for reasons related to analog beamforming: instead of reserving only one slot per beam, it is possible to allocate more than one beam to symbols within the same slot period.

Another advantage of mini-slot implementation is about transmissions in unlicensed spectra when a node has to sense the channel before starting delivery. In such scenarios, it would require a certain amount of time from the instant when the channel is sensed idle and the next slot boundary, hence the solution is to start transmission immediately on the ongoing slot.

The drawback is that probably, mini-slots will not be supported by all devices.

2.4 Structure in frequency domain

Due to the scalable numerology, NR bandwidth channels may be quite large, up to 400 MHz. For this reason, NR devices cannot be required to cover the entire spectrum and, as a consequence, they will be able to handle only a part of the carrier. Furthermore, the part of the carrier received by a

specific equipment may not be centred on the carrier frequency, resulting in a scenario in which each device operates on a subcarrier centred on different frequency. This enforces the deployment of DC subcarriers (*Direct Current subcarriers*) that do not carry useful data but instead they are used by each device to discover the centre of the bands where transmission takes place (Figure 2.3).

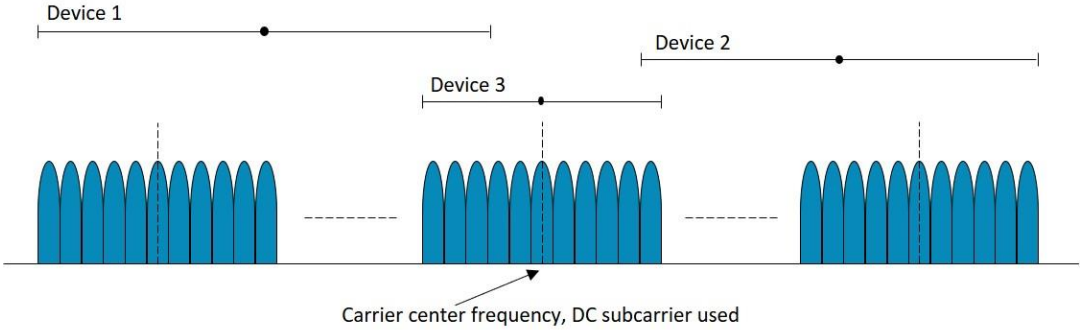


Figure 2.3 *DC subcarrier in NR*

The elementary unit defined in frequency domain is referred to as *resource element* [7] (Figure 2.4).

A resource element is long one OFDM symbol upon one subcarrier in frequency and it is denoted by the pair $(i,j)_{p,\mu}$, where i is the frequency index and j represents the symbol time index, p is the antenna port, μ is the scaling factor.

A set of 12 consecutive subcarriers along one OFDM symbol is then called *resource block* (*RB*).

In contrast to what was implemented for LTE, in NR the resource block must cover necessarily one single OFDM symbol due to the flexibility of subcarrier spacing otherwise it would be nontrivial to guarantee alignment among channels in time domain.

The alignment in frequency domain is maintained, indeed one resource block holding a subcarrier spacing of $2\Delta f$ covers the interval of frequencies occupied by two resource blocks at Δf.

The aligning among resource blocks at different scaling factors is maintained through the definition of *resource grids*.

A *resource grid* consists of $N_{grid,y}^{size,\mu} \times N_{sc}^{RB}$ subcarriers in the frequency domain and $N_{symb}^{subframe,\mu}$OFDM symbols in the time domain, where $N_{grid,y}^{size,\mu}$ is the specific carrier bandwidth

adopted for scaling factor μ on the channel with direction y that may be either *downlink* or *uplink*, while $N_{sc}^{RB} = 12$ is the number of subcarriers per resource block.

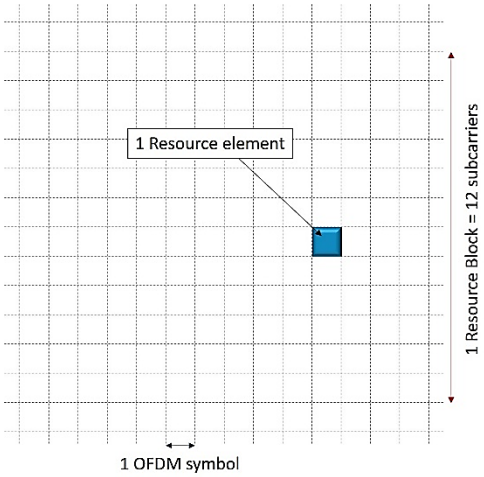


Figure 2.4 *Resource grid and resource element in 5G*

For each scaling factor and antenna port, a specific resource grid is defined.

In order to indicate the resource block within the carrier where the transmission takes place it is used a reference point called as *point A* jointly with two kinds of resource blocks: *common resource block* and *physical resource block.*

Point A is the subcarrier 0 of common resource block 0 for every scaling factor, it is exploited to map physical resource block to the corresponding common resource block.

Common resource blocks are defined in such a way that, given μ, the center of their subcarrier 0 is the reference point A.

If n_{CRB}^{μ} is the index of common resource block, then it holds that

$$n_{CRB}^{\mu} = \left\lfloor \frac{i}{N_{sc}^{RB}} \right\rfloor$$

where i is the subcarrier index relatively to point A.

Physical resource blocks are labelled through a number between 0 and $N_{BWP,b}^{size}$, where b is the number of the bandwidth part.

The mapping between common and physical resource blocks is

$$n_{CRB} = n_{PRB} + N_{BWP,b}^{start}$$

with n_{CRB} being the number of common resource block, n_{PRB} the number of physical resource block and $N_{BWP,b}^{start}$ the initial common resource block of the bandwidth part with respect to the common resource block 0.

As an example, it may be considered the scenario illustrated in Figure 2.5 where for the subcarrier spacing Δf the physical resource block 0 is mapped to the m^{th} common resource block with respect to point A.

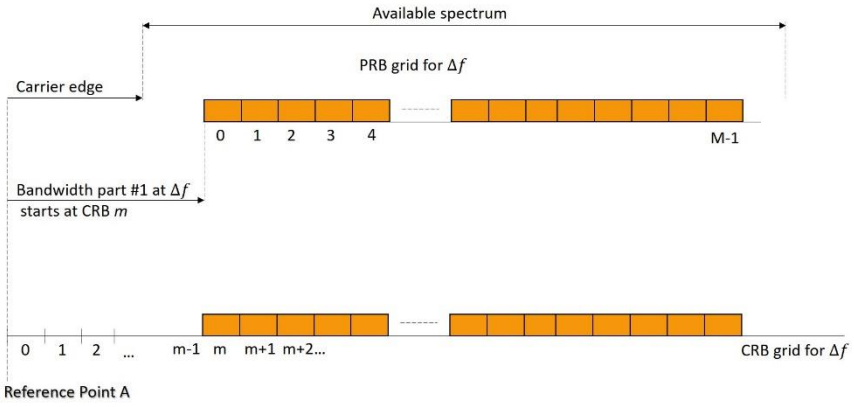


Figure 2.5 *Common and physical resource blocks*

2.5 Bandwidth part

According to specifications in [8], a bandwidth part (BWP) is a set of contiguous resources blocks for a fixed numerology, it is described through two parameters: $N_{BWP,b}^{start,\mu}$ as the starting resource block of the b^{th} bandwidth part and $N_{BWP,b}^{size,\mu}$ being the size of the bandwidth part itself.

Bandwidth parts are a fundamental tool that has been introduced in NR since devices are not capable of receiving the entire carrier bandwidth. Therefore, Bandwidth parts identify the PRBs where each user is able to handle the communication. Furthermore, each user is designed in such a way to receive only one numerology per time.

Each device can be assigned up to four bandwidth parts for downlink and some other four for the uplink. When a user has to initiate a connection, it receives from the PBCH the initial bandwidth part for the downlink, while the PDCCH signals the initial bandwidth part for the uplink.

Among the bandwidth parts allocated for a cell, there is one defined as *active downlink bandwidth part* at a specific time instant. The same hold for the uplink case.

2.6 Carrier Aggregation

In NR, transmissions may take place in parallel over different carriers making the available bandwidth wider and data-rates higher. The latter technique is commonly defined as *carrier aggregation* [3] and it may be implemented among carriers contiguous or non-contiguous in frequency domain, within the same frequency band (*Intraband aggregation*) or belonging to different frequency bands (*Interband aggregation*).

It is allowed to aggregate at most 16 carriers with different duplex schemes.

Usually carriers are referred to as *cell*, among the aggregated carriers one of them is the *primary cell* (PCell) and it is the first cell that a device detects and connects to. Afterword secondary cells may be activated if they are needed.

The number of cells aggregated for uplink is not necessarily equal to cells aggregated for downlink direction.

2.7 Modulation schemes

The modulation schemes supported on NR channels are [9]:

- *QPSK*
- *16-QAM*
- *64-QAM*
- *256-QAM*
- $\frac{\pi}{2}$-*BPSK* (only in uplink for low PAPR and more efficient power amplifier at low data-rates).

Each of the upper schemes is preferred to the others depending on the usage case.

2.8 OFDM waveforms

3GPP agreed on selecting *Orthogonal Frequency Division Multiplexing* (OFDM) as adopted waveforms for 5G New Radio.

OFDM is a peculiar variant of the *Frequency Division Multiplexing* (FDM) in which data stream is mapped onto different subcarriers within the available bandwidth, allowing multiple transmissions at the same time but on different sub-bands [10].

The key property of OFDM stands in the orthogonality among sub-carriers, also called *tones,* so that signals sent over different sub-bands do not interfere to each other. Moreover, each subcarrier is modulated with a modulation scheme at low symbol rate independently from the other tones.

OFDM waveforms are obtained through *Inverse Fast Fourier Transform* (IFFT) that generates OFDM symbols allocated within time slots.

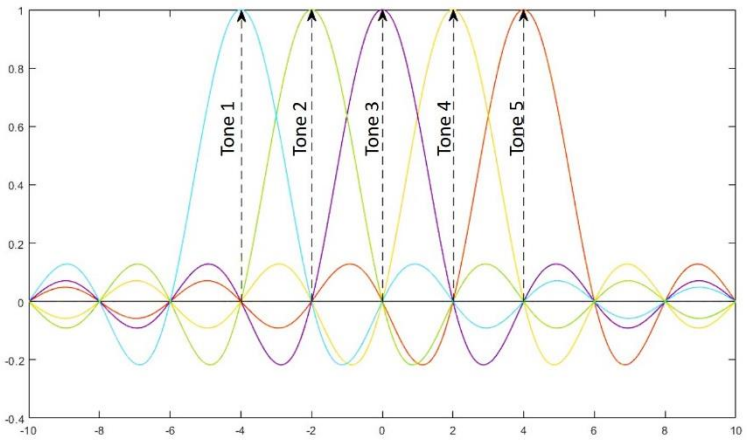


Figure 2.6 *OFDM waveforms*

The version of OFDM adopted in NR is the *Cyclic prefix*-OFDM (CP-OFDM): between two adjacent OFDM symbols a guard interval is inserted on which the cyclic prefix is transmitted. In principle, guard intervals are brief empty portions of time between slots that do not carry payload, having the aim of isolating adjacent transmissions so to prevent *Inter-symbol Interference* (ISI). In the specific case of CP-OFDM, guard time is not void but it is filled with the last portion of the OFDM symbol, denoted as cyclic prefix. CP-OFDM enables the use of cyclic convolution that is more efficient than usual convolution since preserves the initial part of samples useful for example in filtering operations.

The main advantages of OFDM waveforms are:

- Robustness against ISI
- Robustness against *Co-Channel Interference* (CCI)
- Ease of implementation through IFFT/FFT
- Increase of the spectral efficiency

On the other side, OFMD has a relevant issue involving high *peak-to-average-power ratio* (PAPR).

In 5G New Radio, OFDM is deployed both for uplink and downlink directions.

For uplink transmission, it is supported *DFT-precoded* OFDM implying higher power-amplifier efficiency.

DFT-precoding is implemented in order to spread the symbol energy among tones so to make it as uniform as possible. As drawback, DFT-precoding require complex operations at the receiver in the case of spatial multiplexing.

2.9 Transmission techniques

NR has been designed in order to support both schemes *Time Division Duplex* (TDD) and *Frequency Division Duplex* (FDD) [3].

In *Time Division Duplex* transmissions in uplink and downlink are allocated on the same subcarrier but on different time slots so to guarantee the nonoverlapping in time domain.

It is also possible a dynamic version of TDD enabling dynamic allocation of uplink or downlink channels to slots.

In TDD scheme it is necessary to reserve a guard period in order to separate downlink from uplink and to guarantee a transient period to switch from a direction to the other. The length of the guard period is proportional to the cell size, hence it increases as the distance from the base station gets higher.

In *Frequency Division Duplex* uplink and downlink transmissions happen simultaneously in time domain while are assigned different subcarriers, hence uplink and downlink channels are distinguished within the entire available bandwidth.

The FDD scheme require filtering operations in order to guarantee the separation between UL ad DL sub-bands.

A NR device is not intended to support necessarily full-duplex, for example low-cost equipmentmay support only half-duplex links, meaning that simultaneous transmission in both directions are not allowed. In these latter cases, even if FDD is supported, it happens that uplink and downlink are separated even in time domain.

2.10 Physical Channels list

3GPP has established a set of NR physical channels both for the uplink and downlink direction [11].

By definition, a downlink channel is a set of resources allocated for the transmission of information to a UE.

The downlink channels defined for NR systems are:

- *Physical Downlink Shared Channel* (PDSCH), responsible for the delivery of unicast payload but also Random Access Message and paging messages;
- *Physical Broadcast Channel* (PBCH), carrying management information when a user accesses the network;
- *Physical Downlink Control Channel* (PDCCH), handling control messages for scheduling and coordination among transmissions on PDSCH.

Uplink channels are instead the radio link which carries information from a UE.

Uplink channels allocated in NR are:

- *Physical Uplink Shared Channel* (PUSCH), as the PDSCH but in the opposite direction;
- *Physical Uplink Control Channel* (PUCCH), handling feedback messages from user to the gNB about channel estimation, hybrid-ARQ acks and resources requests;
- *Physical Random Access Channel* (PRACH), for random access.

Most relevant physical signals handled on Downlink channels are the following:

- *Primary Synchronization Signal* (PSS), as the first sequence detected by a user as soon as it accesses the network and useful for the synchronization of the device on the carrier frequency of the network itself;
- *Secondary Synchronization Signal* (SSS), being necessary for a device in order to find out the Physical Cell Identity (PCI) after synchronization;
- *Demodulation Reference Signal* (DM-RS), for coherent demodulation of symbols on PDSCH.
- *Channel State Information Reference Signal* (CSI-RS), enabling the channel estimation at the user end but also carrying information about frequency and time synchronization and modulation scheme;
- *Tracking Reference Signal* (TRS), for the compensation of time and frequency misalignments.

The main physical signals supported on Uplink channels are:

- *Sounding Reference Signal* (SRS), enabling the gNB to estimate the uplink channel on different sub-carriers on the basis of the measurements sent by the UE;
- *Phase Tracking reference signal* (PT-RS), for the compensation of phase-shift especially at higher frequencies;
- *Demodulation Reference Signal* (DM-RS), as for the downlink case, for coherent demodulation of symbols on PUSCH.

3. The Physical Downlink Control Channel (PDCCH)

Beyond a broad outline of the physical layer in 5G radio access technology, this document is intended to focus on the *Physical Downlink Control Channel*, commonly referred to as the acronym PDCCH.

The existence of a control channel is fundamental to the management of transmissions over a wireless communication system and, according to 3GPP specifications, two control channels are expected to be implemented in NR: one for downlink and one for uplink direction, respectively PDCCH and PUCCH.

As mentioned previously, a description of PDCCH follows in next sections.

3.1 Purposes

The *Physical Downlink Control Channel* is responsible for handling control signalling enabling the correct delivery of payloads over the PDSCH.

The control information, carried by the PDCCH, mainly deals with:

- resource allocation for DL-SCH;
- modulation schemes adopted (QPSK, 16-QAM, 32-QAM, 64-QAM, …);
- parameters of coding techniques, such as the *code rate*;
- transport block format (size, structure, fields, …);
- H-ARQ signaling, standing for Hybrid automatic repeat request. It is exploited for requesting the retransmission of a transport block in case of error detection;
- downlink scheduling assignments and grants, for the coordination among transport block belonging to different data stream on different subcarriers;
- power control;
- preemption indication.

The latter two purposes are considered extraordinary signaling, providing additional support capabilities to PDCCH.

3.2 Resource allocation

Being the bandwidths significantly wide, for New Radio it would be quite challenging to span the control channel over the entire spectrum. As consequence, the PDCCH does not cover the full bandwidth but only a portion.

The elementary unit on which the PDCCH is built is the *resource element group* (REG), one REG corresponds to one resource block (i.e. 12 subcarriers) [12].

A set of six REGs compose one *control-channel element* (CCE).

The entire set of resources in frequency and time domain allocated for the PDCCH is the *Control Resource Set* (CORESET).

The CORESET consists of 1,2,4,8 or 16 contiguous CCEs (Figure 3.1), depending on the adopted *aggregation level*. The CORESET is the space over which a user searches for the candidate control channels. Increasing the aggregation level allows to map the same amount of information on more physical resources, thus increasing the robustness of the channel, as will be detailed in Section 5.

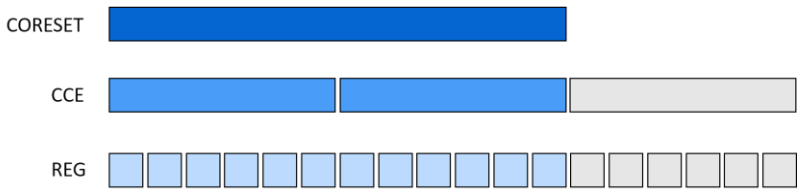


Figure 3.1 *CORESET structure*

Besides the size even the position of the CORESET is variable, indeed it can be located anywhere within the carrier bandwidth and the slot. Typically, the CORESET is at the beginning of the slot but, depending on the application, it may be more suitable to have it at different positions, as in the case of low latency requirements.

In the time domain, the CORESET can be allocated over 1,2 or 3 OFDM symbols. However, the maximum length of the CORESET is 2 OFDM symbols if the PDSCH DM-RS is mapped onto the third OFDM symbol, whereas it is 3 OFDM symbols long if PDSCH DM-RS is placed on the fourth symbol.

The size of the CORESET is not a dynamical parameter: once that the size of the CORESET is set, then it cannot be changed. The static assignment of resources avoids the necessity of the deployment of a further channel to signal the current position and configuration of the PDCCH, as the PCFICH did in LTE.

Another parameter related to the configuration of the CORESET is the mapping between REGs within a CCE that can be *interleaved* or *non-interleaved* (Figure 3.2), the mapping is a property specific for a given CORESET. In order to represent the CCE to REG mapping, it is exploited the concept of REG bundle, being a set of REGs having the same precoding features.

In case of non-interleaved mapping, the REG bundle size is six, and actually the CCE is made up of six contiguous resource elements over one OFMD symbol.

If the mapping is interleaved, then several configurations are possible:

- if the CORESET is defined over one or two OFDM symbols, the REG bundle size may be two or six;
- if the CORESET is defined over three OFDM symbols, the REG bundle size may be three or six.

The interleaved mapping is suitable for frequency diversity due to the spreading of resources elements in the frequency domain.

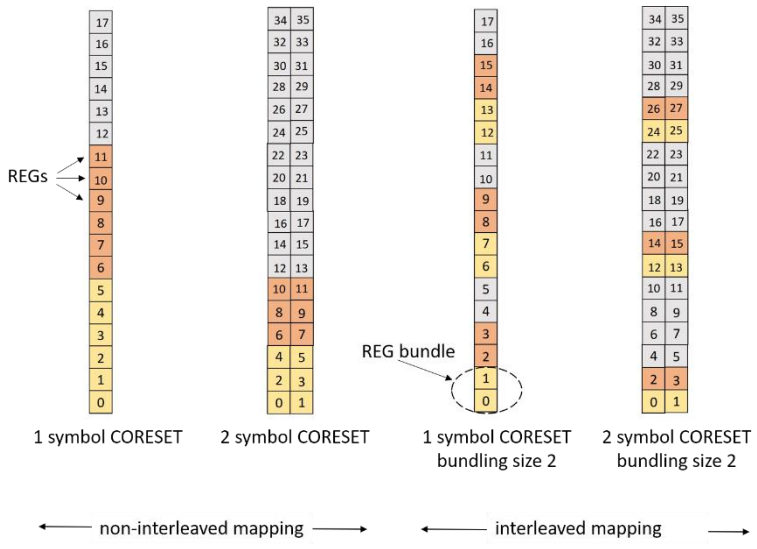


Figure 3.2 *Examples of CCE-to-REG mapping (interleaved and non-interleaved)*

Although the allocation of CORESET is static, the resources assigned to the PDCCH may be reused for the PDSCH. More precisely, some resources overlapping with the CORESET resources may be reserved to serve eventually PDSCH. Then if PDCCH does not use the CORESET resources, the reserved ones for PDSCH are marked as available and it can map data stream on them.

A specific DMRS sequence is expected for the PDCCH and it is obtained by means of a pseudo-random generator.

Furthermore, within the CORESET it is expected one DMRS symbol every four subcarriers.

Multiple CORESETs may be employed to serve a PDCCH, however CORESET 0 is configured by PBCH and, being the first one, is always signalled in the *master information block* (MIB) during the initial part when a UE accesses the network.

3.3 DCI Formats

The transport block transmitted over the physical downlink control channel is denoted as *Downlink Control Information* (DCI) [13].

Two formats are provided: *DCI format 1_0* and *DCI format 1_1*.

DCI format 1_0

It carries control information about the scheduling of the PDSCH across one single cell labeled through an identification number denoted as C-RNTI.

It has a smaller size with respect to the DCI format 1_1 and it represents the fall-back format since terminals and gNB resort to format 1_0 every time they miss some parameters for the proper configuration of a device. Hence possible use cases of DCI format 1_0 are, for instance:

- when a UE detects a base station and attempts to access the network;
- in case of errors over the channel;
- in case of low overhead requirements.

The structure of format 1_0 is illustrated in Table 3-1, where it can be found each field and its respective size in bits.

Field	Size (bits)
DCI format Identifier	1
Frequency domain resource assignment	[X]
Time domain resource assignment	4
VRB-to-PRB mapping	1
Modulation and Coding scheme	5
New data indicator	1
Redundancy version	2
HARQ process number	4
Downlink assignment index (DAI)	2
PUCCH resource indicator	3
PDSCH-to-HARQ feedback timing indicator	3

Table 3-1 *DCI Format 1_0*

DCI format Identifier, indicating whether the DCI is exploited for downlink allocations or uplink grants. It is set to 1 when it is a DL DCI.

Frequency domain resource assignment, signaling the set of resource blocks within a frequency band over which the PDSCH is scheduled.

In principle the size of this field depends on the *allocation type* adopted but in DCI format 1_0 it is possible only *allocation type 1*. According to allocation type 1, a set of contiguously resource blocks within a bandwidth part is assigned to PDSCH. The assignment is inserted in the "frequency domain resource assignment" field through a parameter called *Resource Indication Value* (RIV).

In the RIV, two further parameters are encapsulated:

- the starting virtual resource block (RB_{start})
- the length of contiguously allocated RBs (L_{RBs}).

Then the RIV is computed as

$$RIV = \begin{cases} N_{BWP}^{size}(L_{RBs} - 1) + RB_{start} & if\ (L_{RBs} - 1) \leq \lfloor N_{BWP}^{size}/2 \rfloor \\ N_{BWP}^{size}\left(N_{BWP}^{size} - L_{RBs} + 1\right) + \left(N_{BWP}^{size} - 1 - RB_{start}\right) & otherwise \end{cases}$$

where $L_{RBs} \geq 1$ and $L_{RBs} \leq N_{BWP}^{size} - RB_{start}$.

Being that $RIV_{max} = \frac{N_{BWP}^{size}(N_{BWP}^{size}+1)}{2}$, the maximum size of the Frequency domain resource assignment is $[X] = \left\lceil \log_2 \frac{N_{BWP}^{size}(N_{BWP}^{size}+1)}{2} \right\rceil$ bits.

Time domain resource assignment, indicating the OFDM symbols within a slot on which is allocated the PDSCH.

VRB-to-PRB mapping, signaling whether the mapping between Virtual resource block and physical resource block is interleaved or non-interleaved.

Modulation and Coding scheme, specifying the modulation scheme, code rate and transport block size.

New Data Indicator, indicating whether the codeword is being transmitted for the first time or it is a re-transmission.

Redundancy version, providing the portion of redundancy added by the coding process.

HARQ process number, related to the adopted hybrid-ARQ process.

Downlink assignment index (DAI), indicating the amount of scheduled downlink transmissions for a specific UE.

PUCCH resource indicator, signaling the resources allocated for the PUCCH.

PDSCH-to-HARQ feedback timing indicator, providing the reference time on which HARQ feedback is sent for the downlink stream received.

DCI format 1_1

This format provides a more complete configuration of a device.

The structure of format 1_1 is illustrated in Table 3-2 and it is an extended version of format 1_0 with some additional fields.

Field	Size (bits)
Carrier Indicator	0 or 3
DCI format Identifier	1
Bandwidth part indicator	0,1 or 2
Frequency domain resource assignment	[X]
Time domain resource assignment	4
VRB-to-PRB mapping	1
PRB bundling size indicator	0 or 1
Rate Matching indicator	0,1 or 2
ZP CSI-RS trigger	0,1 or 2
Transport Block 1	8
Transport Block 2	8
HARQ process number	4
Downlink assignment index (DAI)	2
PUCCH resource indicator	3
PDSCH-to-HARQ feedback timing indicator	3
Antenna Port(s)	4, 5 or 6
Transmission configuration indication	0 or 3
SRS request	2 or 3
CBG transmission information (CBGTI)	0, 2, 4, 6 or 8
CBG flushing out information (CGBFI)	0 or 1
DMRS sequence initialization	1

Table 3-2 *DCI Format 1_1*

Beyond the already illustrated fields for format 1_0, the main fields introduced in format 1_1 are described in the following.

Carrier Indicator, signaling the carrier on which PDSCH is scheduled in case of multicarrier configuration.

Bandwidth Part Indicator, representing one of the four allocated bandwidth parts.

Frequency domain resource assignment, being the amount of resources allocated in frequency for PDSCH.

Even in DCI format 1_1 the size of this field depends on the adopted allocation type, that, in this case, can be type 1, 0 or both. Thus, three possible cases may occur:

- if allocation type 1 is chosen, then the same considerations hold as for DCI format 1_0;
- in case of resource allocation 0, the RBG assignment is represented through a bitmap. Hence the "frequency domain resource assignment" field will contain an array of N_{RBG} elements, each one set to 1 or 0 depending if the corresponding resource block is allocated or not. Hence, $[X] = N_{RBG}$ bits;
- if both allocation type 0 and 1 are adopted, then [X] is the maximum value among the sizes at point (a) and (b).

Transport Block 1, Transport Block 2, carrying information about modulation, coding, transport block size.

Antenna Port(s), indicating the antenna port reserved for PDSCH.

SRS request, signaling the request for a Sounding Reference signal transmission.

DMRS sequence initialization, representing which of the possible configuration for DMRS is adopted.

3.4 Search Space

The concept of search space [3] has been introduced in NR since the terminal cannot know the adopted DCI format a priori. The DCI contains information about the index of CCEs belonging to the CORESET, hence until the UE is not able to decode the DCI, the PDCCH position will be unknown. The only feature that the terminal is aware of is the range (search space) where the PDCCH is placed and as consequence the UE will attempt decoding several possible DCI formats in the search space until it will find the CORESET matching.

This procedure based on successive attempts with limited knowledge of information is commonly referred to as *blind detection.*

The search space must be properly sized depending on several aspects.

Firstly, it would be desirable to have a quite small search space, hence the number of candidate CORESETs should be limited. A restricted set of CORESETs is suitable also from the complexity point of view: the smaller is the search space the lower is the complexity required at the terminal side. From the other side, having a reduced search space impacts negatively on the scheduling flexibility. Therefore, the size of search spaces is a trade-off between these factors.

More than one search space can be defined for a single device due to the *aggregation level* that may be 1,2,4,8 or 16, and there is a search space configured for each of them.

In principle, blind decoding attempts to find the PDCCH on search spaces with all possible aggregation levels. In reality, it tries the most likely aggregation levels since, for example, a small sized cell cannot be configured with aggregation level 16.

Furthermore, more than one search space may use the same CORESET as well as multiple CORESETs can be assigned to a device.

The matching between the right PDCCH and the candidate one is found if the CRC in DCI checks.

Two possible search spaces are configured: the *UE-specific search space* and the *common search space.*

UE-specific search space

It is dedicated for a specific terminal. A UE may be configured with more than one specific search spaces over which the user tries to detect the PDDCH using the C-RNTI, being the temporary identifier of a UE in a cell.

Common search space

It is a search space shared by a set of devices and carrying control information that are destined to all of them reducing overhead and complexity. The common search space must be known to all UEs.

DCI exchanged on common search spaces have a set of predefined RNTIs different from the C-RNTI for the specific case. Among possible RNTIs, there is SI-RNTI for scheduling management, P-RNTI for paging, INT-RNTI for preemption signals.

3.5 PDCCH block chain

Operations performed on PDCCH, allowing the delivery of DCI from the gNB to UE end point [3], are represented in the block chain in Figure 3.3.

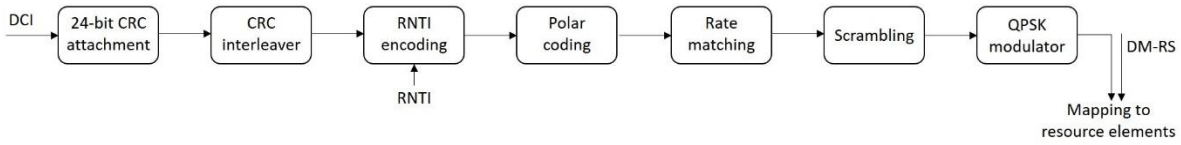


Figure 3.3 *PDCCH block chain*

In the following, the most relevant modules are described more in detail.

CRC attachment

After being generated, the DCI passes through this module that has the aim of computing the CRC and append it on the DCI itself.

The *Cyclic Redundancy Check* is a sequence of bits employed for the error detection at the receiver side. In principle, the CRC is computed through a shift register that takes as input a string and gives as output the corresponding CRC sequence of length L.

In the specific case of the PDCCH, the input string is the DCI and the length chosen for the CRC is L=24 [14].

The logic circuit manipulates the DCI through trivial operations, based on some XOR computations that are determined according to a specific generator polynomial.

Several generator polynomials are defined in the standards of 5G physical layer [14] , for the PDCCH it is selected g_{CRC24C}.

$$g_{CRC24C}(D) = [D^{24} + D^{23} + D^{21} + D^{20} + D^{17} + D^{15} + D^{13} + D^{12} + D^{8} + D^{4} + D^{2} + D + 1]$$

The output is a sequence b of length $M = A + L$, where A is the length of the DCI.

If the DCI is represented through the sequence a and if the CRC is indicated with the sequence p , then the output sequence b is given by:

$$b_j = \begin{cases} a_j & for \;\; j = 0,1, \dots, A-1 \\ p_{k-A} & for \;\; j = A, A+1, \dots, A+L-1 \end{cases}$$

CRC interleaver

The interleaver simply changes the order of the sequence b in order to shuffle payload and CRC bits. However, a certain number of CRC bits is concentrated at the end of the resulting sequence since it will be scrambled with RNTI in the following module.

RNTI encoding

The *Radio Network Temporary Identifier* (RNTI) is an identification number that is exploited in order to identify a connection between a specific UE and a specific eNB.

There are several types of RNTI depending on the kind of messages exchanged and depending on the endpoints involved in the link, for example P-RNTI is for paging messages while SI-RNTI is for broadcast of system information.

Being $\bar{x}$ the RNTI sequence of length H, then the last H bits of the sequence b are summed modulo 2 with the RNTI bits and replaced by the resulting bits.

The receiver will check the correctness of the last part (CRC+RNTI) of the sequence received in order to detect whether errors have occurred or the payload was destined to another user.

This is a trick that enables encapsulating both information for error detection and address of the destination in one single sequence. As consequence, in 5G New Radio there is no distinction between a corrupted data stream and a sequence intended for another destination, indeed they are equally discarded.

Polar encoding

3GPP has introduced an innovative technique for encoding/decoding named by polar encoding.

Polar encoding and decoding are based on a recursive structure with quite low complexity.

The algorithm implemented for decoding is based on *successive cancellation* and *list decoding* in order to reduce the complexity and improve performances. Further improvements are possible through the Successive Cancellation List Decoder with CRC.

Section 3.6 describes more in detail the key features of polar encoding and decoding and their correspondent implementation.

Rate matching

Rate matching is performed in order to match control stream of bits to the actual number of available bits on the PDCCH, depending on the amount of subcarriers allocated for control signaling.

The available number of bits on the physical channel are denoted as Raw bits b_{raw} and they are given by

$$bits_{raw} = Agg_{lvl} * CCE_{size} * N_{Sub}^{PRB} * (1 - DMRS_{rate}) * Q_m$$

where:

- Agg_{lvl} is the aggregation level;
- CCE_{size} is the size of each CCE. In NR a CCE is made up of 6 REGs;
- N_{Sub}^{PRB} is the number of subcarriers per PRB;
- $DMRS_{rate}$ refers to the fact that a DMRS is sent every four subcarriers, so in a PRB $^3/_{12} = {}^1/_4$ subcarriers are occupied by DMRS while $1 - {}^1/_4 = {}^3/_4$ of PRB is available for the transmission of the DCI.
- Q_m is the cardinality of the adopted modulation.

Before performing rate matching, the coded sequence is divided into sub-blocks that are processes by the sub-block interleaver, as it is stated in [14] .

Being d the sequence of coded bits of length N, the main steps are the following:

for $n = 0\ to\ N - 1$

$i = \lfloor 32n/N \rfloor$

$J(n) = P(i) \times (N/32) + mod(n, N/32)$

$y_n = d_{J(n)}$

end for

The obtained sequence is then processed by the rate matching.

The rate matching has the purpose of transforming the coded bits of length N into a sequence of length $E = b_{raw}$.

Two possible scenarios may happen:

- if $E \geq N$ the interleaved sequence is repeated until all the raw bits are filled. This is the repetition method.
- if $E \leq N$:
  - If $K/E \leq 7/16$, the last $N - E$ part of interleaved bits is put on Raw bits.
    Notice that K is the length of sequence before encoding. This is the puncturing method.
  - If $K/E \geq 7/16$, the first part of interleaved sequence is put on the Raw bits. This is the shortening method.

Scrambling

This module re-arranges the input sequence according to the following rule [15]

$$\check{b}(i) = \big(b(i) + c(i)\big) mod 2$$

where

- $b(i)$ is the sequence of length b_{Raw} at the output of the rate-matching;
- $c(i)$ is the scrambling sequence, that depends on parameters of the channel;
- $\check{b}(i)$ is the scrambled sequence.

The scrambling sequence, as stated in [9], is obtained by combining sequences x_1 and x_2.

$x_1(0) = 1, x_1(n) = 0, n = 1,2, \dots ,30.$

$x_2$is the binary version of the number $c_{init} = (n_{RNTI}\ 2^{16} + n_{ID}) mod\ 2^{31}$ [15], where:

- n_{RNTI} is the RNTI;
- n_{ID} is the ID of the cell.

Modulator

As mentioned before, the modulation scheme adopted for the PDCCH transmissions is the Quadrature Phase-Shift Keying (QPSK).

The rule that maps bits to symbol is

$$d(i) = \frac{1}{\sqrt{2}}\left[\left(1 - 2b(2i)\right) + j(1 - 2b(2i+1))\right]$$

where d is the symbol sequence and b is the bit stream at the output of the scrambler [15].

3.6 Polar Codes

Polar codes fall under the category of the error correcting and capacity-achieving codes meaning that given a B-DMC (*Binary-input Discrete Memoryless Channel*) their property is to reach the highest achievable rate of the channel when the input letters have equal probability.

Polar codes have been presented for the first time in 2009 by Erdal Arikan in [16].

Before the introduction of polar codes, Shannon proved the existence of capacity-achieving codes only through the random-coding method, but in 2009 Arikan presented an explicit construction of such codes.

The implementation of polar codes is based on the principle of "*Channel polarization*", i.e. starting from a single channel it is possible to derive N copies of the channel among which some of them are completely noiseless while some others are completely noisy.

Beyond the channel polarization, there are specific operations involved for the encoding and decoding process of polar codes.

In order to better explain how channel polarization works it's necessary to describe the contest in which Arikan implemented them and the corresponding notation.

Given a *binary-input discrete memoryless channel* (B-DMC) W, it's possible to define:

- $(W: \mathcal{X} \to \mathcal{Y})$: a generic B-DMC with X as input alphabet, Y as output alphabet and $W(y|x)$ as the transition probabilities, where $\mathcal{X}$ *is* $\{0,1\}$ always, $\mathcal{Y}$ and $W(y|x)$ can be arbitrary.
- $(W^N: \mathcal{X}^N \to \mathcal{Y}^N)$: the resulting channel derived from N uses of W, where $W(y_1^N | x_1^N) = \prod_{i=1}^{N} W(y_i|x_i)$

- $I(W)$: the symmetric capacity, that is the highest rate achievable if the input letters have same probability. It is given by

$$I(W) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} W(y|x) \log \frac{W(y|x)}{\frac{1}{2}W(y|0) + \frac{1}{2}W(y|1)}$$

$I(W)$ equals the Shannon capacity when the channel is a symmetric channel. The most common examples of binary channels are BSC and BEC.

BSC stands for *Binary Symmetric Channel* where $\mathcal{Y} \in [0,1]$, $W(0|0) = W(1|1)$ and $W(1|0) = W(1|0)$.

BEC stands for *Binary Erasure Channel* and it's a channel where for each $y \in \mathcal{Y}, W(y|0)W(y|1) = 0$ or $W(y|0) = W(y|1)$, with y defined as the erasure symbol.

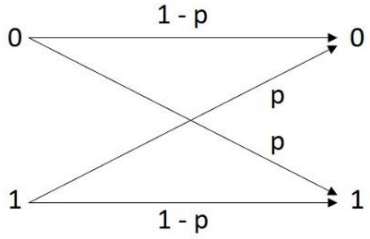


Figure 3.5 *BSC*

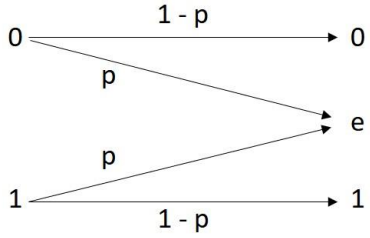


Figure 3.5 *BEC*

If the adopted numerology is in base-2 logarithm, then $I(W) \in [0,1]$.

- $Z(W) \triangleq \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)}$: the *Bhattacharyya parameter*, representing an upper bound on the probability of ML decision error when W is used only once to transmit 1 or 0. Hence $Z(W) \in [0,1]$ and it describe in practice the reliability of the transmission.

It is quite easy to understand that when $Z(W) \approx 0$ then $I(W) \approx 1$ and when $Z(W) \approx 1$ then $I(W) \approx 0$.

Further typical notation involved is:

- X, Y: random variables
- x, y : realizations of respectively $X\ and\ Y$
- P_X: probability measure of X
- $P_{X,Y}$: joint probability measure of X and Y
- $I(X;Y)$, $I(X;Y|Z)$: mutual information and condition mutual information respectively
- a_1^N, a_i^j : row vector of length equal to N and the corresponding sub-vector including elements from the i-th to the j-th
- $A \otimes B$: the Kronecker product between matrices A and B
- $A \otimes B = \begin{bmatrix} A_{11}B & \dots & A_{1n}B \\ \dots & \dots & \dots \\ A_{m1}B & \dots & A_{mn}B \end{bmatrix}$
- $A^{\otimes n} = A \otimes A^{\otimes(n-1)}$, for $n \geq 1$: the Kronecker power of matrix A , assuming $A^{\otimes 0} = [1]$.

Channel Polarization

The very first step in polar codes construction is the manipulation of the B-DMC in a specific way, performed recursively, which gives as result the so-called *channel polarization.*

The main idea behind channel polarization consists in collecting N replicas of the B-DMC (W) and, from this initial set of independent binary channels, through several manipulations, it is extracted a new set still of N binary channels $\{W_N^{(i)}: 1 \leq i \leq N\}$.

The obtained set consists of two possible types of channels: the first is the group of channels for which $I(W) \to 1$, while for those in the second group $I(W) \to 0$.

It is now possible to understand the reason of the term “polarization”. Indeed, among the N channels obtained, some of them will be *good channels* and will be suitable for data transmission, some others are instead *bad channels* where the reliability is too low (near 0) to carry info bit.

The polarization effect is implemented through two phases: *channel combining* and *channel splitting*.

Channel Combining

The goal of this phase is the aggregation of N independent copies of the original channel.

It begins at level 0 from a single B-DMC W referred to as $W_1 \triangleq W$.

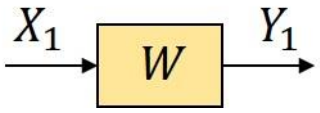


Figure 3.6 *Channel model* $\boldsymbol{W_1}$ *at level 0*

At level 1, W_1 is spliced producing two independent replicas denoted as $W_2: \mathcal{X}^2 \rightarrow \mathcal{Y}^2$.

From the hardware implementation point of view, the combination of the two replicas of W is performed though the mod-2 sum and the obtained diagram is called butterfly pattern as it is illustrated in Figure 3.7:

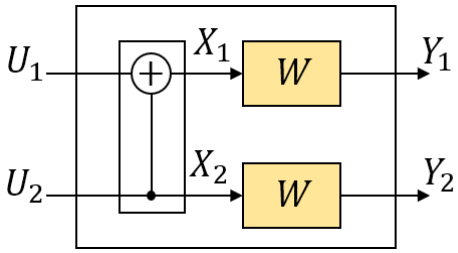


Figure 3.7 *Butterfly pattern of* $\boldsymbol{W_2}$

The transition probabilities of the new channel W_2 will be given by

$$W_2(y_1, y_2|u_1, u_2) = W(y_1|u_1 \oplus u_2)W(y_2|u_2) ,$$

where the vector u_1^N represents the information bits.

At level 2, two identical and independent copies of W_2 are combined resulting in Figure 3.8

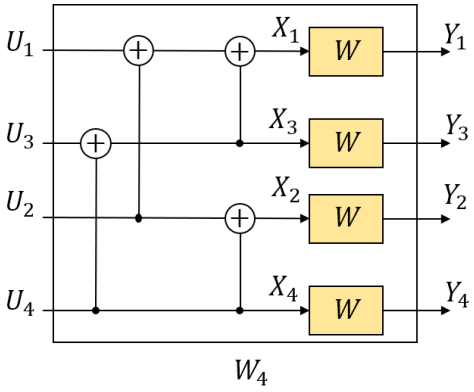


Figure 3.8 *Butterfly pattern of* $\boldsymbol{W_4}$

Hence, it is obtained the channel $W_4: \mathcal{X}^4 \rightarrow \mathcal{Y}^4$ an its transition probability is

$$W_4(y_1^4|u_1^4) = W_2(y_1^2|u_1 \oplus u_2, u_3 \oplus u_4)W_2\,(y_3^4|u_2, u_4).$$

The previous illustration of the channel W_4 can be represented in a more practical way in terms of implementation, as in Figure 3.9:

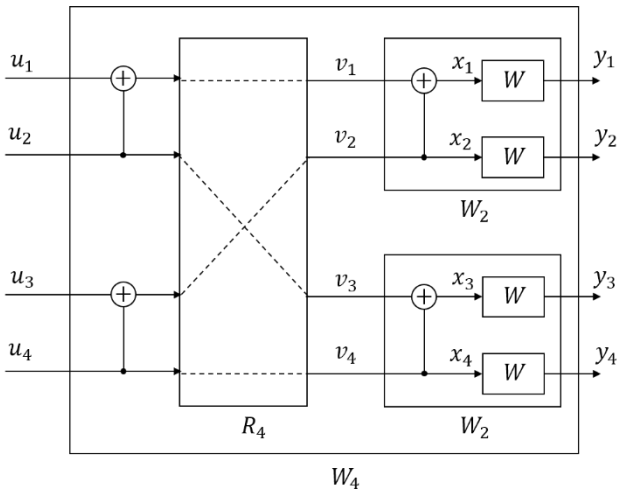


Figure 3.9 *Alternative representation for* $\boldsymbol{W_4}$

This diagram highlights how the information letters u_1^N propagates from the input of channel W_4 to the input of the first level channels W_1.

The module R_4 has the only purpose of mapping the input bits u_1^N to the input v_i^N of the channels W_2 at level 1.

Then, a furthermore manipulation has to be applied in order to reach the input of channels W at level 0 and it simply consists in two sums mod-2.

The overall result of the processes described above can be obtained through a single operation that exploits the matrix $G_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$, therefore $x_1^4 = u_1^4 G_4$.

Exploiting the matrix G_4, it is found out $W_4(y_1^4|u_1^4) = W^4(y_1^4|u_1^4 G_4)$.

It can be shown that the previous steps can be generalized, obtaining $G_N = B_N F^{\otimes N}$ for any $N = 2^n, n \geq 0$, where B_N is the bit-reversal permutation matrix and $F \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

Thus, the channel combining can be reduced to the general case of $N = 2^n$, where the channel W_N is obtained recursively from two channels $W_{N/2}$ (Figure 3.10).

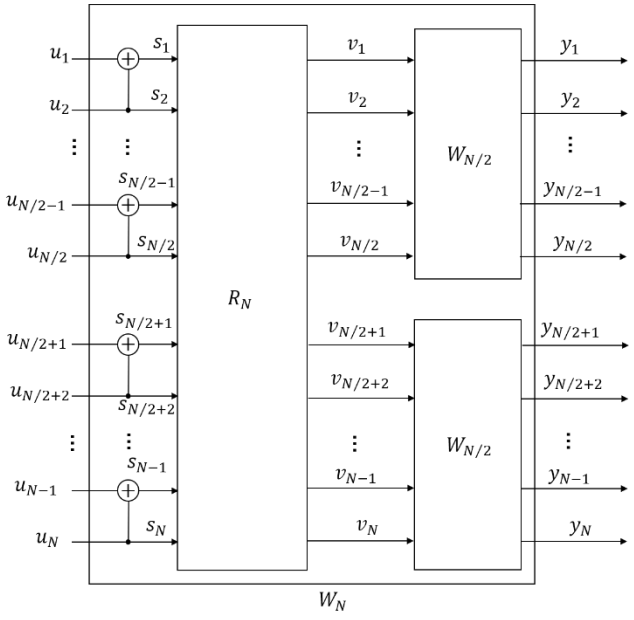


Figure 3.10 *General structure for a channel* $\boldsymbol{W_N}$

The corresponding transition probability is then $W_N(y_1^N|u_1^N) = W^N(y_1^N|u_1^N G_N)$ for all $y_1^N \in Y^N$ and $u_1^N \in X^N$.

The vector channel $W_N: X^N \to Y^N$ represents the channel resulting from the synthetization of N independent channels and preserving their capacity.

Channel Splitting

In the splitting phase the equivalent channel W_N is split into a new set of N channels rearranging the original capacity among them so to satisfy Theorem 1 in [16].

Theorem 1: For any B-DMC, the channels $\left\{W_N^{(i)}\right\}$ polarize in the sense that, for any fixed $\delta \in (0,1)$, as N goes to infinity through powers of two, the fraction of indices $i \in \{0,1,\dots,N\}$, for which $I\left(W_N^{(i)}\right) \in (1-\delta, 1]$ goes to $I(W)$ and the fraction for which $I\left(W_N^{(i)}\right) \in [0, \delta)$, goes to $1 - I(W)$.

In a formal way, the new set of channels are defined as $W_N^{(i)}: X \to Y^N \times X^{i-1}, 1 \le i \le N$ with correspondent transition probability

$$W_N^{(i)}\left(y_1^N, u_1^{i-1} \middle| u_i\right) \triangleq \sum_{u_{i+1}^N \in X^{N-i}} \frac{1}{2^{N-1}} W_N(y_1^N|u_1^N)$$

where (y_1^N, u_1^{i-1}) is the output of $W_N^{(i)}$ and u_i is its input sequence.

In practice, at step i it is found that the i^{th} channel $W_N^{(i)}$ describes the probability on the entire vector y_1^N and past channel inputs u_1^{i-1} on the basis of u_i .

Channel Polarization

The global effect of channel polarization, including both combining and splitting phases, can be presented in a more intuitive fashion, exploiting binary trees to highlight how the mutual information is reallocated among the final set of N channels [17].

At the beginning the system consists of one single binary channel W and by combining two independent replicas of it, it is possible to obtain two polarized channels: $(W, W) \to (W^-, W^+)$.

To each one of these two channels will be associated the corresponding capacity I and the Bhattacharyya parameter Z, then one can choose to use any of them for computations.

Leading operations using the capacity, if $I(W)$ denotes the capacity of channel W then $I(W^-)$ and $I(W^+)$ will represent the capacities of W^- and W^+ respectively. The latter operation is the single-step transform.

The powerful property, making the construction of polar codes possible, is that $I(W^-) + I(W^+) = 2I(W)$ and $I(W^-) \leq I(W) \leq I(W^+)$ (with equality only if I(W) equals 0 or 1.), this means that $I(W^-)$ will never equal $I(W^+)$ and in addition among the two channels there will always be one better (W^+) than the other(W^-).

As it is explained in [16], in the case of a BEC channel, it is:

- $I(W^-) = I(W)^2$
- $I(W^+) = 2I(W) - I(W)^2$

Therefore, at level 1 it would be as in Figure 3.11

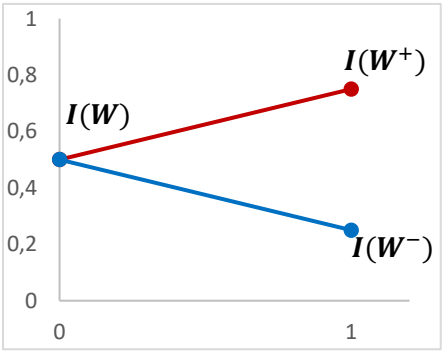


Figure 3.11 *Channel polarization at level 1*

At level 2, it is performed the same transformation to W^- and W^+recursively, as in Figure 3.12:

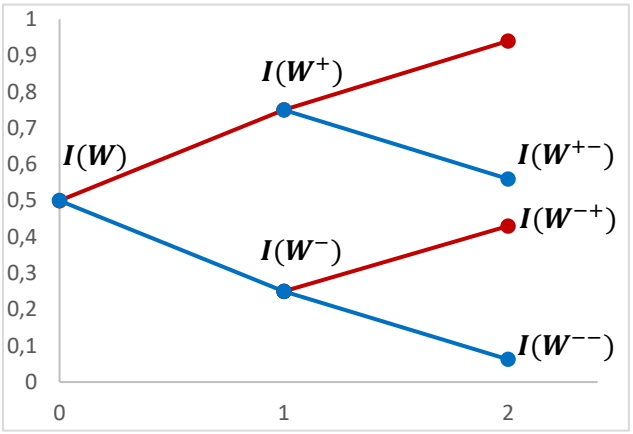


Figure 3.12 *Channel polarization at level 2*

$I(W^-)$ generates $I(W^{--})$and $I(W^{-+})$ while $I(W^+)$ generates $I(W^{+-})$ and $I(W^{++})$, being

- $I(W^{--}) = I(W^-)^2$
- $I(W^{-+}) = 2I(W^-) - I(W^-)^2$
- $I(W^{+-}) = I(W^+)^2$
- $I(W^{++}) = 2I(W^+) - I(W^+)^2$

with $I(W^{--}) \leq I(W^{-+}) \leq I(W^{+-}) \leq I(W^{++})$.

At the n^{th} level, the original channel is transformed into N polarized channels

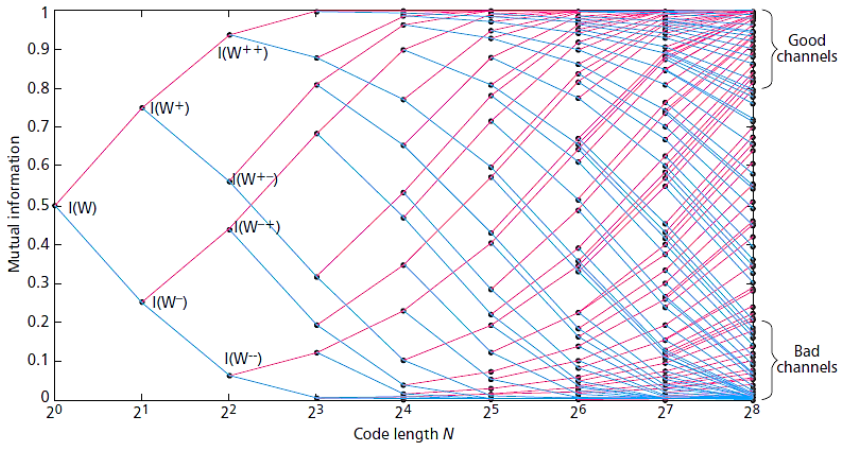


Figure 3.13*Channel polarization at level n*

For N that goes to infinity the polarization effect gets more accentuated, in the sense that the symmetric capacity distribution of good channels tends to get denser around $I(W)$, whereas for bad channels it's more concentrated around $1 - I(W)$.

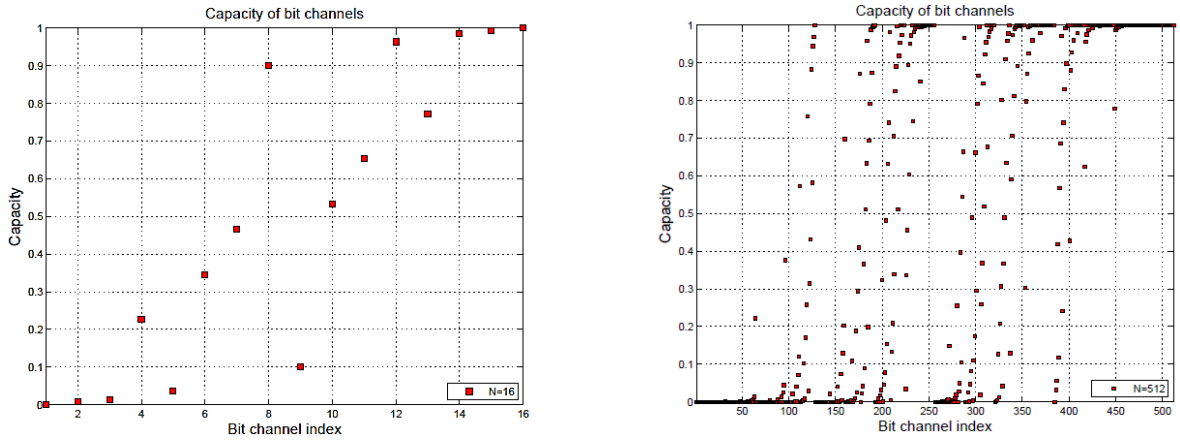


Figure 3.14 *Capacity distribution for* $\boldsymbol{N = 16}$ *and* $\boldsymbol{N = 512}$

Polar coding

By means of polar coding, it is indicated the set of tools that converge to a coding system where each channel $W_N^{(i)}$ can be accessed to send data only if $I(W_N^{(i)}) \approx 1$.

Assuming K as the information sequence at the input of the channel W and N as the coded block length, with $N > K$.

The channel polarization will be performed recursively until N polarized channel will be extracted from the initial binary channel, then the information bits are allocated to the K most reliable channels, while the remaining $N - K$ channels will carry *frozen bits*.

Frozen bits are those bits fixed to a predefined value (typically 0) and they are assigned to bad channels.

Considering for example a BEC(1/2) with N=8, K=4, then the allocation would be as depicted below exploiting both butterfly pattern and binary tree approaches:

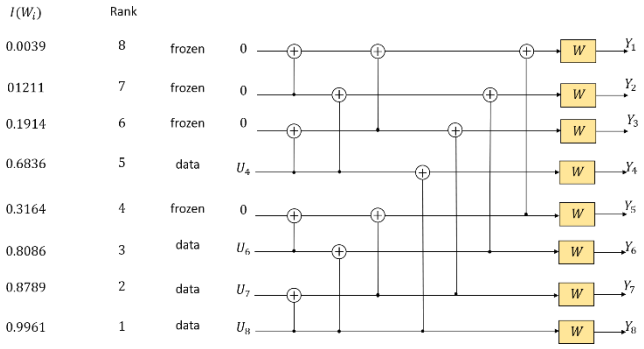


Figure 3.16 *Butterfly pattern for* $\boldsymbol{N = 8, K = 4}$

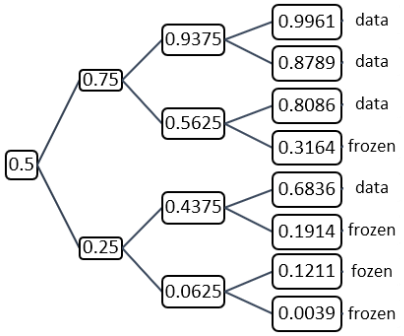


Figure 3.15 *Binary tree for* $\boldsymbol{N = 8, K = 4}$

Therefore, once N and K are set and polarization is performed, the following parameters will be defined:

- u_1^N: the bit sequence at the input of the encoder
- A : the information set being a subset of $\{1, \dots, N\}$ containing the indices of the K most reliable channels on which information bits will be assigned
- A^C: the complement of set A, with elements corresponding to indices of bad channels
- u_A , u_{A^C} : as the array containing only elements of u_1^N with index equal to elements in A and A^C respectively
- G_N: the generator matrix of order N
- $G_N(A)$: the submatrix of G_N containing rows with indices in A

Finally, the coded sequence can be obtained as

$$x_1^N = u_1^N G_N = u_A G_N(A) \oplus u_{A^C} G_N(A^C)$$

Polar coding is defined through the set of parameters (N, K, A, u_{A^C}) denoted as the $G_N - coset\ code$.

A $G_N - coset\ code$ is the class of codes using matrix $G_N(A)$ as generator matrix with the coset determined by the vector $u_{A^C} G_N(A^C)$.

Several observations follow:

- There are no constraints on u_A

- The ratio K/N quantifies the code rate
- Polar codes are channel-specific designs: once that $G_N - coset\ code$ having (N, K, A, u_{A^c}) is fixed, the correspondent polar codes are unique and are valid only for that specific channel.

The complexity for polar encoding is $O(Nlog(N))$.

Polar decoding through Successive Cancellation

Arikan's design for polar codes exploited the technique of the *successive cancellation* for the decoding phase.

The input of the decoder is the sequence y_1^N and assuming A and u_{A^c} known, the aim of the decoder is to compute the estimate $\hat{u}_1^N$ of u_1^N.

Taking into account that the position of frozen bits is known, it is possible to set $\hat{u}_{A^c} = u_{A^c}$, hence the estimation purpose comes down to the set $\hat{u}_A$.

An important feature to be considered is the fact that the butterfly pattern at the encoder introduces a certain correlation between bits at the lower and upper branches. Thus, in the following stages further correlation is introduced so that each coded bit will be affected by all its preceding bits with smaller indices.

This aspect justifies the deployment of the term "successive cancellation", as it is necessary to remove the influence caused by previous bits.

The algorithm implemented in [16] assumes the received vector y_1^N to contain LLRs (Log-Likelihood Ratio).

The decoding process still exploits the butterfly pattern and it involves the propagation of two types of messages: LLRs from right to left and bits from left to right [18] (Figure 3.17).

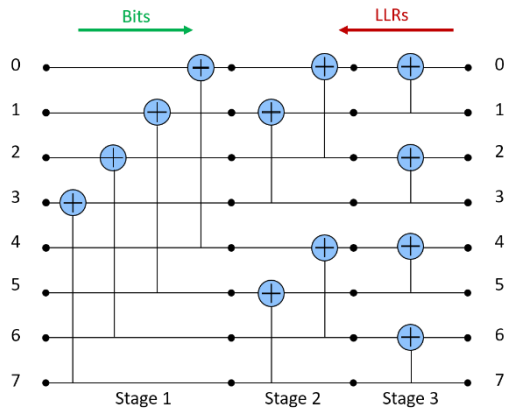


Figure 3.17 *Decoding process rule*

At the beginning only the LLRs at the output of Stage 3 are available, hence they propagate toward left.

Bits are recovered only when LLRs propagation gets to the root nodes (nodes on the left end of stage 1). Therefore, through hard detection on LLR at a certain root node it is decided whether the corresponding bit is a 0 or 1.

The rule to be applied for the updating of LLR is as follows [17]:

$$L_{i,j} = \begin{cases} 2\tanh^{-1}\left[\tanh\left(\frac{L_{i+1,j}}{2}\right)\cdot\tanh\left(\frac{L_{i+1,j+2^{i-1}}}{2}\right)\right] & \left\lfloor\frac{j-1}{2^{i-1}}\right\rfloor mod2 = 0 \\ \left(1-2s_{i,j-2^{i-1}}\right)\left(L_{i+1,j-2^{i-1}}\right)+L_{i+1,j} & otherwise \end{cases}$$

By approximation:

$$(1)\ L_{i,j} \approx \begin{cases} sgn\left(L_{i+1,j}\right)\cdot sgn\left(L_{i+1,j+2^{i-1}}\right)\cdot\min\left(\left|L_{i+1,j}\right|,\left|L_{i+1,j+2^{i-1}}\right|\right) & \left\lfloor\frac{j-1}{2^{i-1}}\right\rfloor mod2 = 0 \quad (a) \\ \left(1-2s_{i,j-2^{i-1}}\right)\left(L_{i+1,j-2^{i-1}}\right)+L_{i+1,j} & otherwise \quad (b) \end{cases}$$

where

- i: i-th stage
- j: j-th level

- $L_{i,j}$: LLR value at node corresponding to i-th stage and j-th level
- $s_{i,j}$: Bit at i-th stage and j-th level recovered by hard detection

Practically, the rule for updating LLRs at nodes in previous stage $L_{i,j}$ depends whether the considered node is at lower or upper branch.

In the trivial the case with i=1 and j=1,2, it would be:

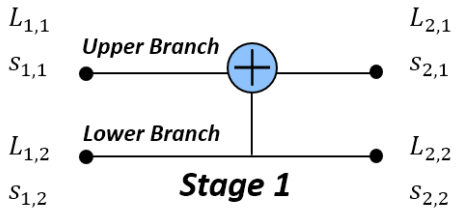


Figure 3.18 *LLR and bits (i=1, j=1,2)*

At the very starting point the decoder receives as input $L_2 = (L_{2,1}, L_{2,2})$.

Then it proceeds in broadcasting L_2 towards the previous stage nodes, so exploiting (1) the correspondent LLRs are recovered.

As mentioned before, it holds:

- if the LLR under assessment is the one at upper branch ($L_{1,1}$), it applies (1a)
- if the desired LLR is the one at lower branch ($L_{1,2}$), it applies (1b).

Going deeply into details of the working principle of the algorithm, it is useful to describe it step by step. To better highlight all the involved operations, it is more convenient to examine the case with i=1,2 and j= (1,…,4). The trellis under assessment is in Figure 3.19.

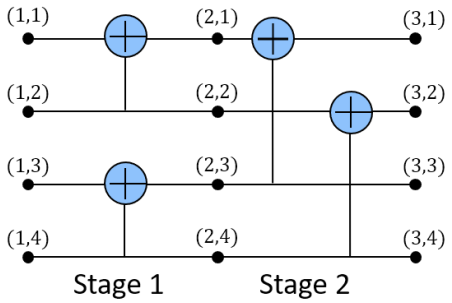


Figure 3.19 *Butterfly pattern i=1,2 j=1,2,3,4*

According to adopted notation, each pair (i,j) denotes the node at the beginning of the i^{th} stage and at the j^{th} level.

To each pair (i,j) is then associated a LLR and a bit value, $L_{i,j}$ and $s_{i,j}$ respectively.

Therefore, when the decoder gets the LLRs as L_3, their manipulation is carried out as it follows:

1. $L_{2,1}$ and $L_{2,2}$ are estimated separately exploiting (1a). For example, $L_{2,1}$ will derive from the knowledge of $L_{3,1}$ and $L_{3,3}$
2. $L_{1,1}$is estimated putting $L_{2,1}$ and $L_{2,2}$ in the (1a)
3. by means of hard detection on $L_{1,1}$, it is performed the ML decision on the corresponding bit $s_{1,1}$, in general it holds:

$$(2) \qquad s_{i,j} = \begin{cases} 0 & if \quad L_{i,j} \geq 0 \\ 1 & otherwise \end{cases}$$

4. the knowledge of $s_{1,1}$, $L_{2,1}$, $L_{2,2}$ makes possible the update of $L_{1,2}$ exploiting (1b)
5. $s_{1,2}$ can now be recovered after the decision taken on $L_{1,2}$ according to (2)
6. $s_{1,1}$ and $s_{1,2}$ are broadcasted towards right, and according to (3) it is possible to find bits at nodes (2,1) and (2,2). Respectively $s_{2,1}$ is given by (3a) and $s_{2,2}$ by (3b)

$$(3) \qquad s_{i+1,j} = \begin{cases} s_{i,j} \oplus s_{i,j+2^{i-1}} & if \quad \left\lfloor \frac{j-1}{2^{i-1}} \right\rfloor mod2 = 0 \quad (a) \\ s_{i,j} & otherwise \quad (b) \end{cases}$$

7. expression (1b) allows to recover likelihoods at nodes (2,3) and (2,4) separately:

- $L_{2,3}$ is derived from the knowledge of the set $(s_{2,1}, L_{3,1}, L_{3,3})$
- $L_{2,4}$ is derived from the knowledge of the set $(s_{2,2}, L_{3,2}, L_{3,4})$

8. $L_{1,3}$ can be recovered putting $L_{2,3}$ and $L_{2,4}$ in (1a)
9. Hard detection on $L_{1,3}$ gives the estimated bit $s_{1,3}$
10. From the set $(s_{1,3}, L_{2,3}, L_{2,4})$, the expression (1b) provides $L_{1,4}$
11. Finally, the ML decision taken on $L_{1,4}$ provides bit $s_{1,4}$.

An important remark is about the way frozen bits are treated. The decoder has the knowledge of u_{A^c}, hence when the algorithm is run and it gets to a root node with index belonging to u_{A^c}, it sets automatically the correspondent bit to the value of frozen bit.

Decoding through successive cancellation has a low complexity being $O(NlogN)$.

The main drawback of SC decoding is related to the fact that its error correction performance is poor when the code block length N is quite low, indeed polar codes are able to reach the exact capacity of the channel only for N that goes to infinity. Therefore, a successful attempt to improve the performances of the decoder was proposed in [19].

Successive Cancellation List Decoder

Successive Cancellation List decoder is an enhanced version of the SC decoder described previously.

The key point on which SCL is founded is the mechanism that leads to discover each bit [19]. In the algorithm of SC decoder, each time that at a root node LLR is computed then it is transformed into a bit by means of hard detection. Once that a decision is taken on a bit, it is broadcasted and there is no possibility of changing it in the following steps of the decoding process and in addition it will be influent on the following bits decisions. Hence, although it has been declared that polar codes have excellent performances when N goes to infinite, it is quite intuitive that in real applications it does not still hold because in those cases N is finite and too small to exploit the efficiency of polarization. As consequence, it is quite likely that the SC decoder could introduce errors and propagate them.

Furthermore, a second issue in the SC decoder algorithm is related on how frozen bits are treated: the decoder knows the values of future frozen bits but it does not take any advantage of it.

The improvement enabled through the list decoder stands in the fact that each time that a LLR is found at a root node, instead of deciding whether it is 0 or 1, it inspects both possibilities.

The detection is implemented as a binary search tree where two possible paths leaves from each node: one going to 0 and the other to 1, each of them will be associated to a likelihood.

Each level of the binary tree corresponds to a new unfrozen bit $\hat{u}_i$ to estimate, its final height will be equal to N.

The set of branches from the root of the tree to a certain leaf is called "path", therefore a path is determined by the decision taken at each level.

The binary tree can be easily built, in fact when an unfrozen bit $\hat{u}_{i+1}$ has to be decoded, each node at i^{th} level is split into two, generating two possible paths: one appending a bit set to 0 and the other appending a bit set to 1 to the previous path.

A fundamental parameter in SCL decoder is the *list size* L representing the maximum number of candidate paths to be further examined, pruning is performed in case the paths generated at a certain stage are greater than L

To give a more practical explanation it is useful to consider an example where $L = 4$.

- As soon as the first unfrozen bit $\hat{u}_1$ is reached, according to the decoding order, it can be both 0 or 1: two possible paths are generated denoted as ${\hat{u}_0}^1$.
  ${\hat{u}_0}^1 = \{0\}, \{1\}$.

Figure 3.20 *Binary tree at level 0*

- When the second unfrozen bit $\hat{u}_2$ has to be detected it is considered again the possibility to be either 0 or 1, for all the two possible paths leading the first bits: four possible paths ${\hat{u}_0}^2$ are now generated. They're all kept in memory since the maximum number of likely paths is 4.
  ${\hat{u}_0}^2 = \{00\}, \{01\}, \{10\}, \{11\}$.

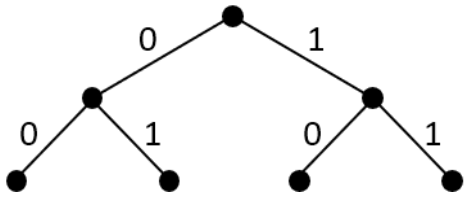


Figure 3.21 *Binary tree at 1st level*

- the following bit to be decoded is $\hat{u}_3$, all options are considered for each one of the 4 nodes at the previous level generating eight paths of the kind $\hat{u}_0{}^3$
  $\hat{u}_0{}^3 = \{000\}, \{010\}, \{100\}, \{110\}, \{001\}, \{011\}, \{101\}, \{111\}$.

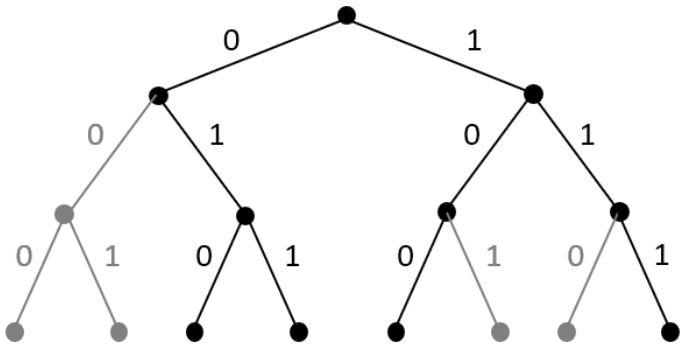


Figure 3.22 *Binary tree at 2nd level*

Notice that now the generated paths are greater than L=4, so pruning is able to remove the worst (least probable) paths and it keeps the 4 most likely ones.

- when the algorithm routine gets to $\hat{u}_4$ detection, the splitting mechanism produces again eight paths, hence pruning is needed again

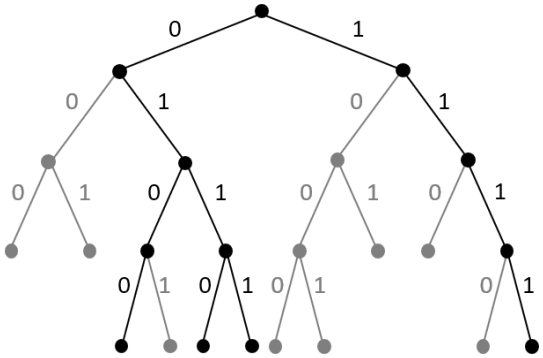


Figure 3.23*Binary tree at 3rd level*

When all the unfrozen bits have been "visited" throughout the algorithm routine, the growth of the search tree stops. At that point the SCL decoder will select the single most likely path.

Successive Cancellation List Decoder has better performances than the simple SC decoder and as L increases error rates get lower and the behaviour gets closer to the ML-decoding.

But the larger is L, the greater is the memory needed as well as the complexity.

Successive Cancellation List Decoder with CRC

A further enhancement involves the use of CRC [19] .

In this this version of the SCL decoder, a small part of the unfrozen bits is reserved for the coding of the CRC bits.

More precisely, if k is the number of the unfrozen bits, then it is set a small constant r and the first k-r bits are assigned to information bits. The last r unfrozen bits will contain the CRC computed on the first k-r bits [20].

The big advantage arises right along the decoding phase as it is explained below.

Assuming the SCL decoder has already found out the most likely L paths leading to $\hat{u}_{k-r}$ (as they are referred to information bits), the next bit to estimate is the first CRC bit.

For simplicity, the subset of CRC-bits in the estimated sequence $\hat{u}$ are denoted as $\hat{u}_{CRC+j}, for\ j = 0,..,r-1$.

Each one of the most likely L paths will contain a possible sequence of k-r bits, ${\hat{u}_0}^{k-r}$ and it is possible to compute the corresponding CRC to each of them.

This is useful in order to predict $\hat{u}_{CRC}$ so that all decision leading to wrong first CRC bit will not be taken into account and paths with incorrect CRC will be removed from the list.

The policy implemented for the selection of paths at each level can be stated as follows:

- if at least one path has a correct CRC, then all paths with corrupted CRC will be excluded from the L paths. Among the survived paths, the most likely will be chosen
- otherwise, the most likely path is selected with an awareness that at least one bit is corrupted and in the hope of reducing the number of wrong bits.

SCL decoders are able to achieve performances significantly close to ML decoding at lengths as short as $n = 2048$.

SCL algorithm is easy to implement, its complexity is $O(L \cdot n \cdot \log n)$, the required memory is $O(L \cdot n)$.

3.7 Beamforming

The physical downlink control channel supports Massive MIMO and beamforming approaches in order to have better performances in terms of throughput and capacity.

Firstly, it's necessary to clarify the meaning of terms Massive MIMO and beamforming and their respective use cases.

MIMO stands for *Multiple Input Multiple Output* and commonly refers to a scenario where more than one antenna, organised into arrays, are exploited at the transmitter and/or at the receiver side. Such implementations are suitable to achieve diversity in order to deal with the fading introduced in

real world environments. Indeed, it can be proved that the channel experienced by each antenna is uncorrelated to channels seen by others. The term massive is then employed in order to highlight the large number of antennas in each array.

By means of *beamforming* it is implied all the procedures that allow to properly set phase and amplitude of each transmitting element so that all the beams combine into a unique beam focused towards a specific direction with the proper narrowness. Beamforming can be performed even at the receiver side, where receiving elements parameters are properly weighted in order to concentrate their reception in a given direction. On the other side, beamforming allows to reduce interference arriving from other directions.

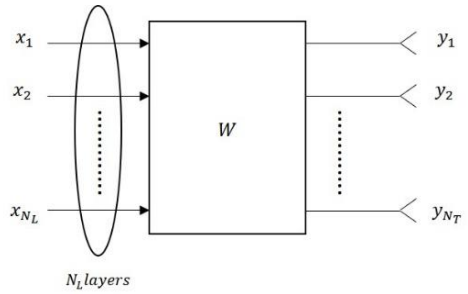
Another use case of MIMO is the so-called *spatial multiplexing*, aiming to transmit data over multiple layers and over the same frequency and time resources, separating the different transmission layers exploiting the spatial characteristics of the channel.

MIMO technology deployment is furthermore motivated by the high frequencies at which 5G New Radio will work [3]. Indeed, it is well known that the higher is the working frequency, the smaller is the area of transmitting/receiving antenna, on the other side it is also true that the directivity is proportional to the size of the antenna. Hence, high frequencies impact negatively on the directivity of a single antenna. It is at this point that MIMO efficiency arises: multiple small radiating elements are put together in an antenna array resulting in a larger size antenna so that high directivity is achievable again.

In view of the above discussion, it is reasonable that beamforming is particularly applied for higher frequencies while, MIMO transmission is exploited for interference reduction and spatial multiplexing at lower frequencies.

In some cases, spatial multiplexing and beamforming can be applied simultaneously.

In Figure 3.24 it is possible to see the scheme of a linear multi-antenna transmission system.



$$\bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_T} \end{bmatrix} = W \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_L} \end{bmatrix} = W\bar{x}$$

Figure 3.24 *General model of multi-antenna transmission*

Where N_L is the number of uncorrelated layers, N_T is the number of transmitting antennas, W is a $N_T \times N_L$ matrix allowing the mapping of the vector $\bar{x}$ on the vector $\bar{y}$.

Multi-antenna processing may be applied in two possible levels within the overall block-chain of the channel:

- in the digital domain, hence before the digital-to-analog conversion at transmitter side;
- in the analog domain, hence after the digital-to-analog conversion, at transmitter side.

For both possible choices, some issues are involved.

Digital beamforming requires a quite high complexity at the transmitter side since it is necessary to have one digital-to-analog converter for each transmitting element, while analog beamforming does not allow FDD transmissions towards terminals located in different directions with respect to the base station, in practice such transmissions must be led over different time intervals.

Multi antenna processing includes all the operations that allows to properly weight each element in the radiating array, this procedure is also called *precoding*, and the matrix W is the *precoding matrix*.

The precoding matrix can be built through the reporting of a set of channel characterizations included in the so-called *Channel State Information* (CSI).

The CSI contains the spatial transfer function between each pair of transmitting and receiving antennas.

CSI acquisition is possible through two possible methods:

- *follow PMI* at gNB with CSI acquisition based on CSI-RS transmitted in downlink
- *eigen-beamforming* at gNB with CSI acquisition based on SRS transmitted in uplink

CSI report contains at least one of the following parameters:

- *Rank Indicator* (RI), being the most appropriate rank according to the device measurements. The rank corresponds to the number of transmission layers N_L in downlink;
- *Precoder-Matrix Indicator* (PMI), being the most appropriate precoder matrix according to the terminal, once the rank has been selected;
- *Channel-Quality Indicator* (CQI), being the most appropriate channel-coding and modulation scheme according to the terminal, once that precoding matrix has been selected.

PMI is reported in the follow PMI approach for the CSI acquisition, where the gNB sends to the terminal CSI-RS, then basing on the received signal the UE perform some measurements of the channel that are exploited to estimate the CSI. The terminal computes the PMI based on the CSI and send it back to gNB. At this point the base station is able to proper design the beamforming.

When PMI is signalled to gNB from the terminal, it is implied that a set of possible precoding matrices should be known at both end-sides and PMI is just an index to select the most suitable among them. The set of precoding matrices is actually defined as precoder codebook and it contains at least one codebook for each combination of N_T antenna ports and N_L layers.

Precoding books have been defined in [21].

Furthermore, the precoding codebook is used mainly by the UE for selecting the PMI, instead the network may apply a codebook different from those in the precoding codebook.

Two possible schemes are provided in [21] for the CSI estimation at the UE end-side:

- *type I CSI*, for cases when a single user is scheduled over a time/frequency resource. Even high-order multiplexing is allowed, achieving transmission on large number of layers;
- *type II CSI*, for MU-MIMO cases, when multiple users are assigned simultaneously over the same resource but with a limited number of transmission layers.

Type I codebooks are simple and their purpose is to focus the transmitted energy towards the target receiver, there is no processing for interference management since it is a receiver skill.

Type II codebooks are instead more complex and contain more informations. Furthermore, such codebooks allow to minimize the interference caused by devices to each other on the same physical resource.

In this document, type I CSI is mainly focused.

Type I CSI codebook

The precoding matrix W is obtained as

$$W = W_1 W_2$$

where W_1 and W_2 being two matrices carrying different information for PMI separately.

It holds that

$$W_1 = \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$$

with $B = [b_0 \ \dots \ b_{L-1}]$ being a matrix containing L neighbour beams b_i, where L is the number of beams per group. The same beam is exploited for the two orthogonal polarizations.

Firstly, it is selected W_1 and it is the equivalent operation of selecting the direction of the beam among the set of all the admitted directions.

By selecting the matrix W_2, it is chosen one of the L neighbours beams and QPSK co-phasing of the two polarizations.

In release 15 of NR, it is supported a set of possible layouts as it is illustrated in Table 3-3

Number of CSI-RS antenna ports, P_{CSI-RS}	(N_1, N_2)	(O_1, O_2)
4	(2,1)	(4,1)
8	(2,2)	(4,4)
	(4,1)	(4,1)
12	(3,2)	(4,4)
	(6,1)	(4,1)
16	(4,2)	(4,4)
	(8,1)	(4,1)

Number of CSI-RS antenna ports, P_{CSI-RS}	(N_1, N_2)	(O_1, O_2)
24	(4,3)	(4,4)
	(6,2)	(4,1)
	(12,1)	(4,4)
32	(4,4)	(4,1)
	(8,2)	(4,4)
	(16,1)	(4,1)

Table 3-3 *Possible layouts for single panel configuration*

(N_1, N_2) represents the antenna ports position in azimuth and elevation respectively, while (O_1, O_2) are the DFT oversampling factors. In total, $N_1 O_1$ possible beams can be generated in azimuth and $N_2 O_2$ possible beams in elevation (Figure 3.25).

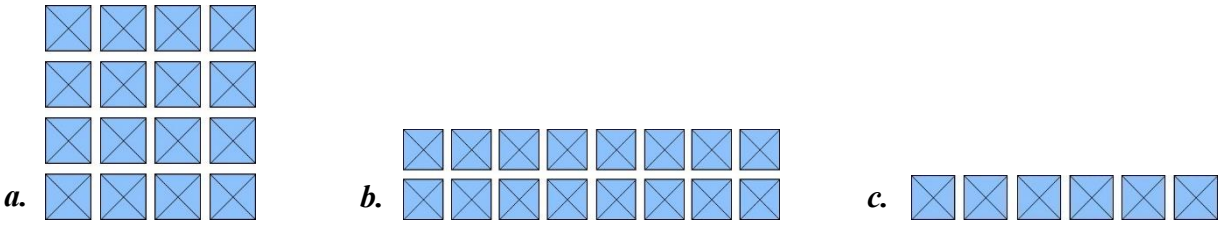


Figure 3.25 *Examples of single panel configurations: a.* $(\boldsymbol{N_1, N_2}) = (\boldsymbol{4, 4})$*; b.* $(\boldsymbol{N_1, N_2}) = (\boldsymbol{8, 2})$*; c.* $(\boldsymbol{N_1, N_2}) = (\boldsymbol{6, 1})$

The precoding matrix is then chosen on the basis of the beamforming vector

$$v_{l,m} = \left[u_m \quad e^{j\frac{2\pi l}{O_1 N_1}} u_m \quad \cdots e^{j\frac{2\pi l(N_1-1)}{O_1 N_1}} u_m\right]^T$$

$v_{l,m}$ is obtained as the Kronecker product between a steering vector in the vertical plane u_m, and one in the horizontal plane v_l, being respectively:

- $u_m = \left[\begin{array}{cccc} 1 & e^{j\frac{2\pi m}{O_2 N_2}} & \cdots e^{j\frac{2\pi m(N_2-1)}{O_2 N_2}} \end{array}\right]$
- $v_l = \left[\begin{array}{cccc} 1 & e^{j\frac{2\pi l}{O_1 N_1}} & \cdots e^{j\frac{2\pi l(N_1-1)}{O_1 N_1}} \end{array}\right]$.

Once l and m are selected, the beam is thus tuned to properly point to the desired direction.

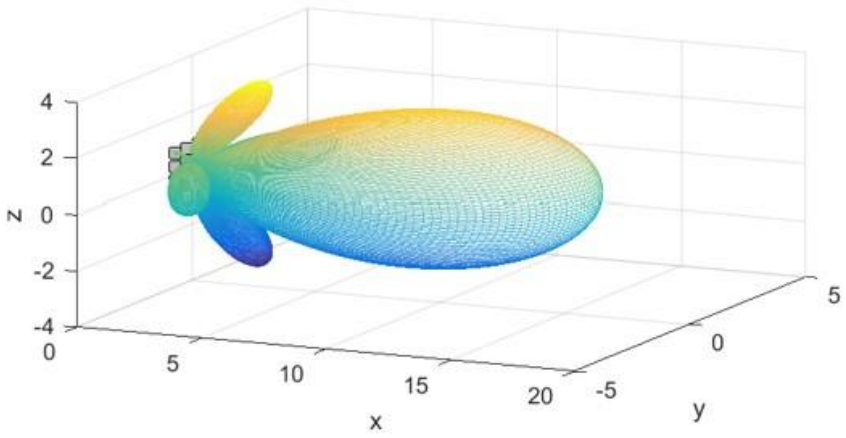


Figure 3.26 *Example of a beam*

Eigen-Beamforming is an alternative approach to Follow PMI for the CSI acquisition. In this method, the CSI estimation is performed at gNB based on SRS transmitted in uplink from the UE. Using SRS the gNB obtains an estimate of the channel matrix H_{SRS}. Beamforming vectors are then obtained computing the eigenvectors of the covariance matrix $R = H_{SRS}^{H} \cdot H_{SRS}$ of the channel. This approach can be followed when a TDD duplexing scheme is used in NR, so that downlink channel can be estimated on the basis of uplink reference signals thanks to the reciprocity of the channel. A follow PMI approach is on the other hand typically used for FDD duplexing schemes, since in this case reciprocity cannot be exploited being uplink and downlink transmission performed on different carrier frequency. In this case the gNB must rely on user measurements and precoding is actually decided by the UE and communicated to the gNB thanks to the PMI.

4. PDCCH simulator

As final goal, this project is aimed to evaluate the performances of the Physical Downlink Control Channel. For this scope it has been implemented a simulator emulating the transmission of a DCI from a gNB to a user terminal over the physical layer.

4.1 Description of the software

The simulator reproduces the entire block chain of a physical channel where each module has been designed according to the specifications released by 3GPP for the processing of the transport block over the PDCCH.

The implementation has been carried out mainly in Matlab except for some functions that for reasons related to computational speed have been implemented in C language.

In the following, it is provided a description of the software for the PDCCH simulator.

The basic structure of the main contains a cascade of functions, each one representing a module in the physical channel.

The simulator starts with a loop on the 'sim_number' parameter in case that more than one simulation with different parameters are launched together.

An inner loop is on the 'SINR_ITER' parameter in order to simulate transmissions for different values of *Signal-to-Noise-plus-Interference Ratio* (SINR).

A further loop is on the 'counter' variable, that counts the simulated time slots, in order to simulate a large number of Transport Block transmissions (in the order of 10^4) for each value of SINR, so that statistically relevant results can be found.

In the following, the most relevant functions for control sequence processing are described.

Figure 4.1 is a capture of the code implemented for the control sequence processing at the transmitter side.

```
for sim_number = 1 : 1
    for SINR_ITER = 1 : length(env.SINR_dB)
        for counter = 0 : env.NTTI-1
            [ tb1 ] = tb_generator(env);
            [ tb4 ] = nr_pdcch_polar_encoder ( env , tb1 );
            [ tb5 ] = nr_pdcch_bit_scrambling(env,tb4);
            [ tb6 ] = nr_pdcch_modulation(env,tb5);
            [ tb8 ] = nr_pdcch_subcarrier_mapping_dl(env,tb7);
            [ tb9 , W] = nr_mimo_processing_tx(env,tb8,H_srs);
            [ tb10 ] = nr_csirs_mapping_dl(env,tb9);
            [ tb11 ] = ofdm_modulation(env,tb9);
            [ tb12 ] = oversample_filter_opt(env,tb11);
```

Figure 4.1 *Functions at the transmitter side*

- *tb_generator.m* generates the DCI as a sequence of random numbers set to 0 or 1
- *nr_pdcch_polar_encoder.m* encapsulates several modules. It performs the CRC attachment, polar encoder, rate matching and interleaver processing.
- *nr_pdcch_bit_scrambling.m* implements the scrambling as it has been described in the section *scrambling* in 3.5.
- *nr_pdcch_modulation.m* maps bits to symbols in the QPSK constellation
- *nr_pdcch_subcarrier_mapping_dl.m* allocates symbols into the corresponding resource elements according to lookup tables built on the basis of specifications in [11]. Several lookup tables are defined each one exploited for the mapping of different type of bits streams. In particular, there is a lookup table for control symbols over the PDCCH, one for DMRS on PDCCH and one for data symbols over the PDSCH.
- *nr_mimo_processing_tx.m* is for the design of beamforming through the proper weighting of elements in the precoding matrix. Different methods are supported for the CSI acquisition, in particular it can be chosen among the *follow PMI* and *eigen-beamforming*, as they are described in section 3.7. Thus, data stream is assigned to different transmitting antennas N_{TX} among N_{LAY} layers (that in the PDCCH case is only one).

- *nr_csirs_mapping_dl.m*, activated in case of CSI acquisition through the follow PMI approach. This function simulates the gNB that inserts CSI reference signals into the resource grid.
- *ofdm_modulation.m* allows to transform the signal from the frequency to the time domain. Hence, each symbol is mapped into an OFDM waveform and properly fills the cyclic prefix.
- *oversample_filter_opt.m* is to simulate the digital-to-analog conversion, it adds samples between symbols (oversampling) and simulates the filtering in the analog domain thus obtaining the signal as it is output from the transmitter antennas.

The signal generated at the transmitter antennas is then input to the MIMO channel, as shown in Figure 4.2.

```
[ tbl3, Hid ] = fading_channel(env,ch,tbl2);
```

Figure 4.2 *Module for the channel implementation*

- *fading_channel.m* emulates the transmission of the signal over the fading channel adding impairments due to diversity. The module can simulate different type of channels, using a Tapped Delay Line (TDL) approach or a Clustered Delay Line (CDL) approach, with different fading profiles according to [5]. A CDL approach allows to capture the spatial characteristics of the channel, and is therefore preferred to better evaluate the behavior of beamforming used at the transmitter.

Figure 4.3 outlines the design of the receiver. Many of the functions are exactly the symmetric version of the corresponding function at the transmitter side. Some other functions are instead implemented only for the receiver side and are described below.

```
            [ tb14 ] = downsample_filter(env,tb13);

            [ tb15 ] = ofdm_demodulation(env,tb14);

            [ tb16 ] = nr_pdcch_pilot_compensation_dl(env,tb15);

            [ tb17 ] = nr_pdcch_channel_estimation_dl(env,tb16);

            [ pmi_A, pmi_B, H_srs, H_csirs ] = nr_fdb_reporting(env,tb15,Hid);

            [ tb18, tb19 ] = nr_pdcch_subcarrier_demapping_dl(env,tb17,tb15);

            [ tb20, tb21 ] = nr_mimo_processing_rx(env,tb18,tb19);

            tb22 = tb20;
            tb23 = tb21;

            [ tb24 , tempsinr ] = symbol_to_bit_demap_pdcch(env, tb22 , tb23);

            [ tb25 ] = nr_pdcch_bit_descrambling(env,tb24);

            [ tb27] = nr_pdcch_polar_decoding (env , tb25, tb1);

            [ bler, raw_ber, dec_ber , thr] = pdcch_performance_calc(env,tb1,tb4,tb27,tb25);

        end
    end
end
```

Figure 4.3 *Functions at the receiver side*

- *nr_pdcch_pilot_compensation_dl.m* performed at the receiver side to extract pilot symbols from the received signal. The position of pilots in the frequency-time grid is found by reading in the corresponding lookup table for DM-RS. Pilots are exploited to evaluate and compensate for the distortion introduced by the channel so that the points of the received constellation can be put back to the right position.
- *nr_pdcch_channel_estimation_dl.m.* One DM-RS symbol is transmitted every 4 subcarriers, thus pilots are only in a restricted number of resource elements. Once that they are extracted it is necessary to perform interpolation in order to recover Demodulation Reference Signal also in the intermediate resource elements. The set of pilots over the entire grid in frequency and time allows to obtain a complete estimate of the channel.
- *nr_fdb_reporting.m* used to generate the feedback information necessary for beamforming. When the follow PMI method is used, this function estimates the index of the precoding matrix (PMI) that should be applied by the gNB based on the channel estimated on CSI-RS, and reports it in uplink. When eigenbeamforming is used, this function estimates the channel as it would be seen at the gNB, based on the SRS signals transmitted in uplink.

- *pdcch_performance_calc.m* has the purpose of computing the performance in terms of:

  - Raw Bit Error Rate (BER_{raw}) as the measure of errors affecting the bits already processed by the encoder. This type of bits is defined Raw since they are the effective bits that cross the channel, after being processed at the transmitter side.

$$BER_{raw} = \frac{bits_{raw,e}}{bits_{raw,tx}}$$

    where $bits_{raw,e}$ is the amount of raw bits received wrong and $bits_{raw,tx}$ is the total number of transmitted raw bits.

  - Decoded Bit Error Rate (BER_{dec}), representing the amount of bits received erroneously after being decoded. The decoder is able to recover some raw bits affected by errors, hence BER_{dec} performances are in general better than BER_{raw}.

$$BER_{dec} = \frac{bits_{dec,e}}{bits_{dec,tx}}$$

    where $bits_{dec,e}$ is the amount of decoded bits received wrong and $bits_{dec,tx}$ is the total number of transmitted bits before being encoded.

  - Block Error Rate (BLER). A block includes the total number of bits sent during one transmission, in this simulator one block corresponds to one DCI transmission.

$$BLER = \frac{TB_e}{TB_{tx}}$$

    where TB_e is the number of received Transport Blocks affected by errors and TB_{tx} is the total number of transmitted Transport Blocks.

    One transport block is considered wrong if it contains at least one wrong bit.

- Throughput being the amount of bits correctly received per time unit

$$Throughput = \frac{bit_c}{T_{tot}}$$

where bit_c is the number of bits received correctly and T_{tot} is the duration of the simulated transmission (i.e. the duration of a slot times the total number of slots simulated).

5. Simulation results

The final goal of this thesis project is the evaluation of PDCCH performances in different scenarios. Each scenario is determined by changing the most relevant parameters that are the aggregation level, the Signal-to-Noise plus Interference Ratio, the DCI format and the approach adopted for the beamforming (follow PMI or eigenvector beamforming).

For each of the scenarios mentioned above, both the real and ideal cases have been simulated.

In the following, several figures report the results of major interest since they allow to draw some conclusions on the innovative aspects of 5G New Radio.

Effect of the aggregation level

The results in Figure 5.1 and Figure 5.2 are relative to the case of ideal channel estimation, eigenvector beamforming and DCI Format 1_1 but similar performances have been found for real scenarios, beamforming with follow PMI and DCI Format 1_0.

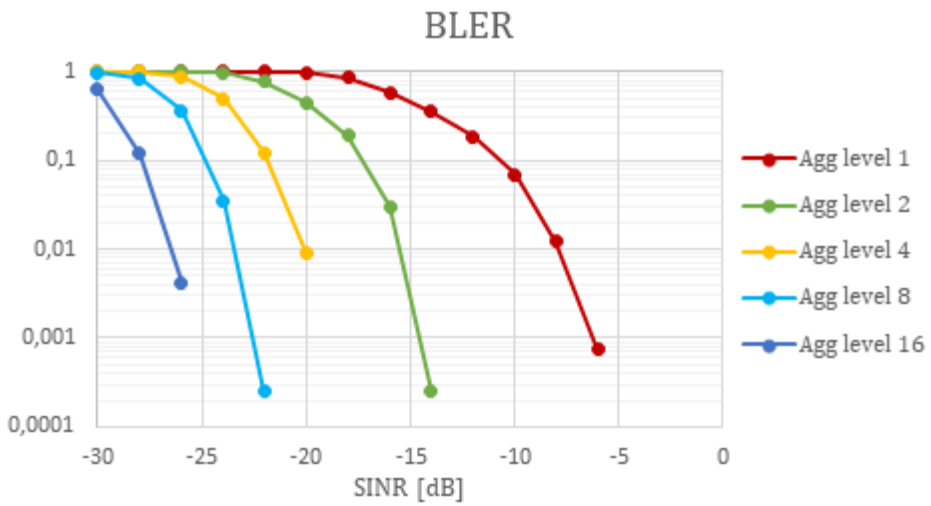


Figure 5.1 *BLER performances for different aggregation levels*

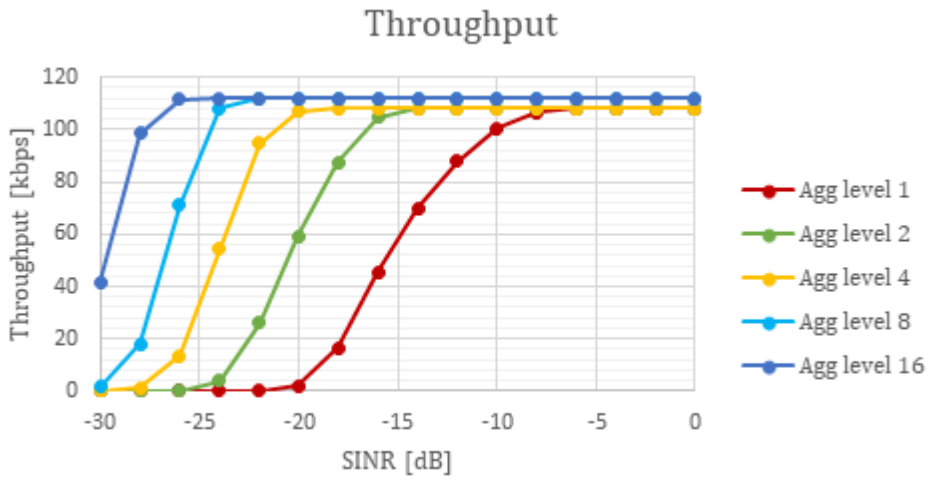


Figure 5.2 *Throughput performances for different aggregation levels*

It can be clearly noticed that for higher values of aggregation level the BLER performances improve as well as the throughput experienced.

In general, this is motivated by the fact that when more physical resources are available, the gain increases too and good performances are achieved at lower SINR.

Going deeply into details, the gain obtained when the aggregation level increases is determined by several effects that are examined below.

In Figure 5.1 there is a gain of $18\ dB$ between the curve corresponding to aggregation level 1 and the one with aggregation level 16. In principle, a doubling of the aggregation level corresponds to a doubling of the available resources, hence the gain experienced passing from aggregation level 1 to 16 would be equal to $10\log_{10}(16) \cong 12\ dB$.

A further improvement leads to a gain around $15\ dB > 12dB$ due to the deployment of coding and in particular polar coding. This aspect can be seen in Figure 5.3 where a simple chain is implemented (Figure 5.4) in order to evaluate the enhancement related to higher aggregation level when polar codes are involved in a transmission over an AWGN channel. Indeed, when more physical resources elements are allocated, the code rate increases. This means that a larger number of parity bits is deployed so that the information bits are more protected and the coding system is able to recover a larger number of corrupted bits.

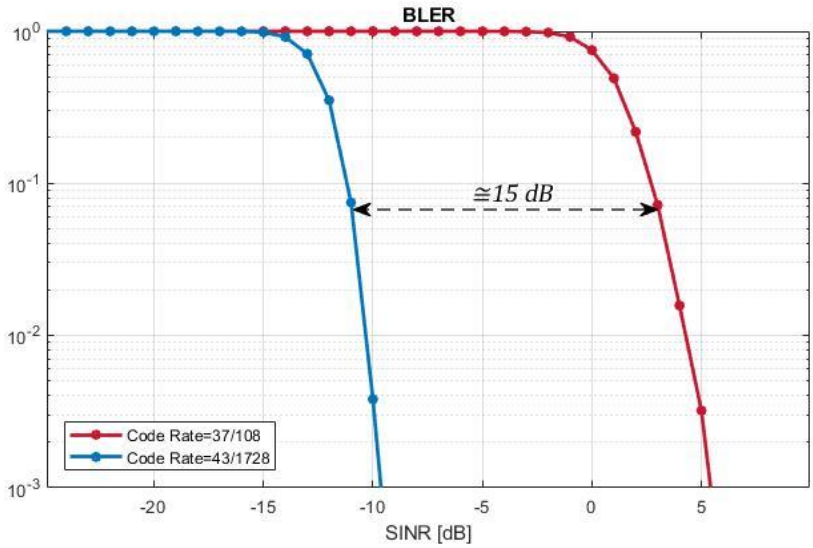


Figure 5.3 *BLER performances for polar encoding/decoding over an AWGN channel*

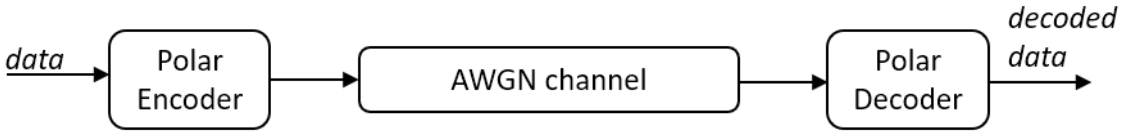


Figure 5.4 *Cascade of polar encoder and decoder and AWGN channel*

Moreover, the PDCCH has been tested for transmissions over a fading channel that is different from an AWGN channel. Indeed, the advantage of a fading channel is that over certain frequencies different copies of the signal combine in a certain way so that the power of the received signal increases, this leads to the so-called *constructive interference*. When the channel bandwidth increases the effect of diversity in frequency is more pronounced. Indeed, in the case of aggregation level 1 the channel bandwidth chosen is equal to 10 MHz, while for the aggregation level 16 it is 80 MHz, thus the channel is averaged over a larger number of resources. As a result, the performances from different aggregation levels gets even better and, at the end, the total gain is about 18 dB .

Effect of polar coding

Figure 5.5 and Figure 5.6 let us appreciate the advantage of the deployment of a coding system that in the particular case of the PDCCH is the polar coding.

The difference between BER_{Raw} and BER_{Dec} has been explained in section 4.1.

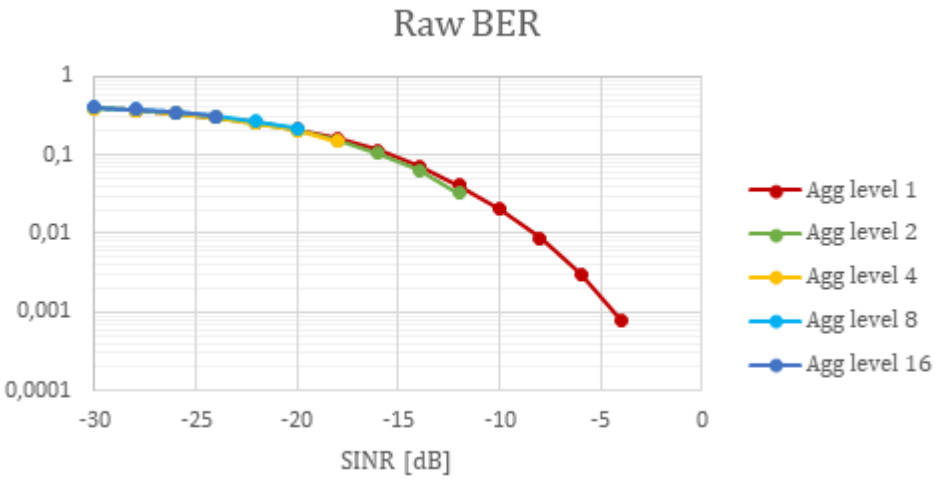


***Figure 5.5 BER_{Raw}** performances for all the aggregation levels*

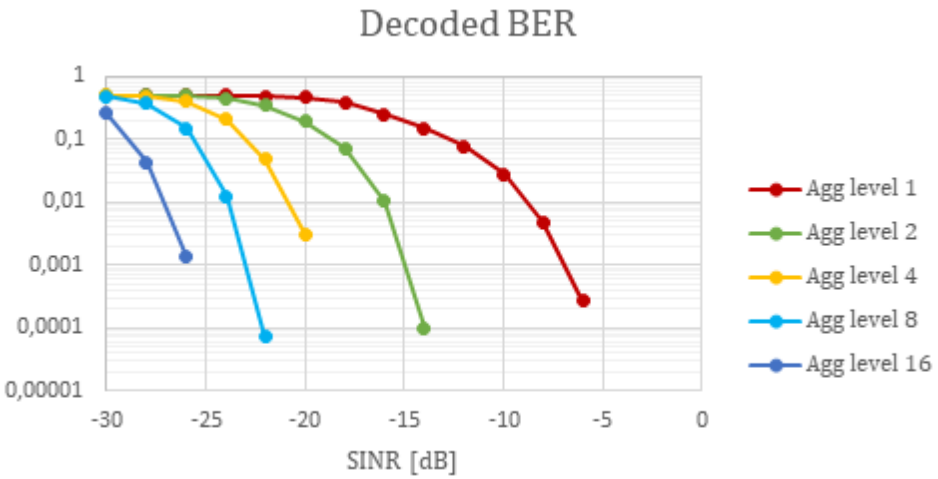


***Figure 5.6 BER_{Dec}** performances for all the aggregation levels*

As it can be noticed in Figure 5.5, there are not substantial differences between curves of BER_{Raw} for different aggregation level. This is justified by the fact that raw bits are not processed by the decoder, hence they are bits simply transmitted over the channel and the increasing of the aggregation level only implies that a larger number of bits is sent. The BER_{Raw} curves are uniquely affected by the statistic that is not affected by the aggregation level.

Figure 5.6, instead, outlines the effect of the decoder and in particular that the more resources are available, the larger is the number of parity bits and the decoding process is able to recover a larger number of bits affected by errors.

Comparison between "follow PMI" and beamforming through eigenvectors

Figure 5.7 reports the behavior of the two different approaches for the management of beamforming: the follow PMI and the eigenvector method. In both cases, it has been simulated the ideal channel estimation case with aggregation level 1 and 16. The ideal condition refers to the use of ideal channel estimation for both channel compensation and ideal pointing of the beam.

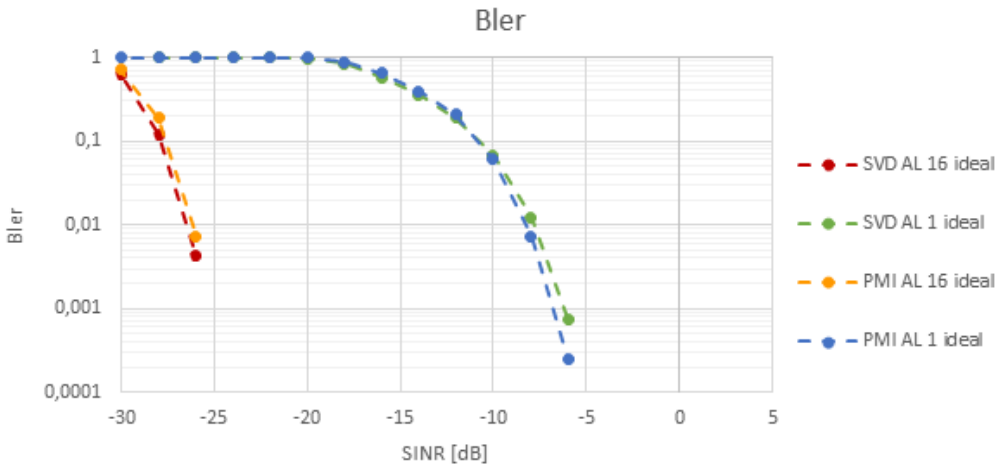


Figure 5.7 *Comparison between follow PMI and eigenvector beamforming*

It can be noticed that results obtained with the follow PMI approach are very similar to the eigenvector case and this can be attributed to the high density of the codebook adopted for the follow PMI; indeed the amount of possible precoding matrices is intensified through oversampling in the spatial dimension.

In the results shown up to now, it is supposed an ideal knowledge of the channel for both channel compensation and beamforming design. In reality, in particular for low Signal to Noise ratio, some impairments overcome along the channel estimation so that the performances in terms of error probabilities and throughput are compromised.

The following sections aim to highlight the dissimilarities observed through the simulations in real channel estimation scenarios.

Follow PMI: ideal vs real channel estimation behavior

Figure 5.8 shows the performances of the follow PMI with aggregation levels 1 and 16. The dashed lines are the same figured in the previous section while the solid lines represent the real scenario where some non-idealities are introduced in the knowledge of the channel.

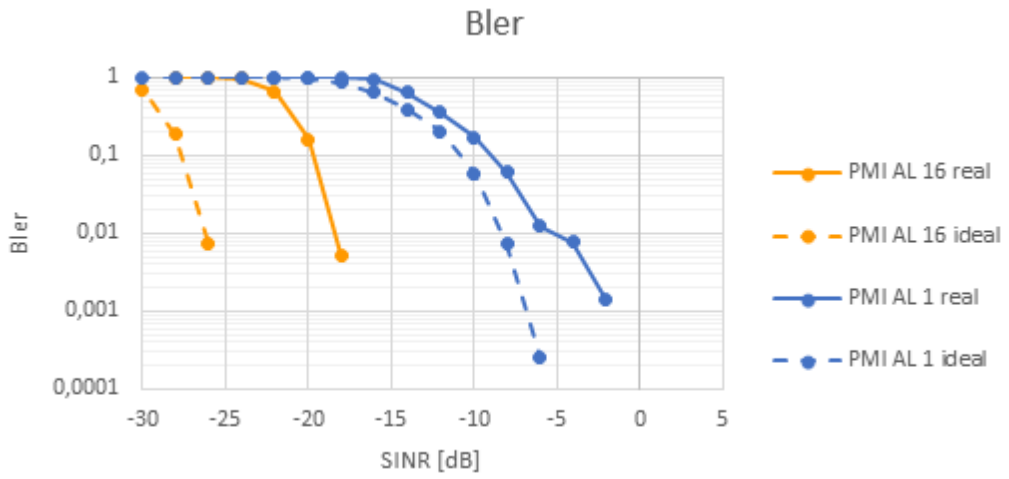


Figure 5.8 *Follow PMI performances in ideal and real scenarios (including non-ideal channel compensation)*

In the case of aggregation level 1, the real curve is quite close to the ideal one, differences can be explained with limitations in the beam focusing and on the channel compensation.

In the case of aggregation level 16 differences become much more relevant.

To understand the main cause of error, it has been tested what happens if the channel compensation is still ideal but the beam focusing is not. In this case, as it can be seen in Figure 5.9, this new curve is still very close to the curve with fully real channel estimation, showing that the main cause of error is actually the beam focusing.

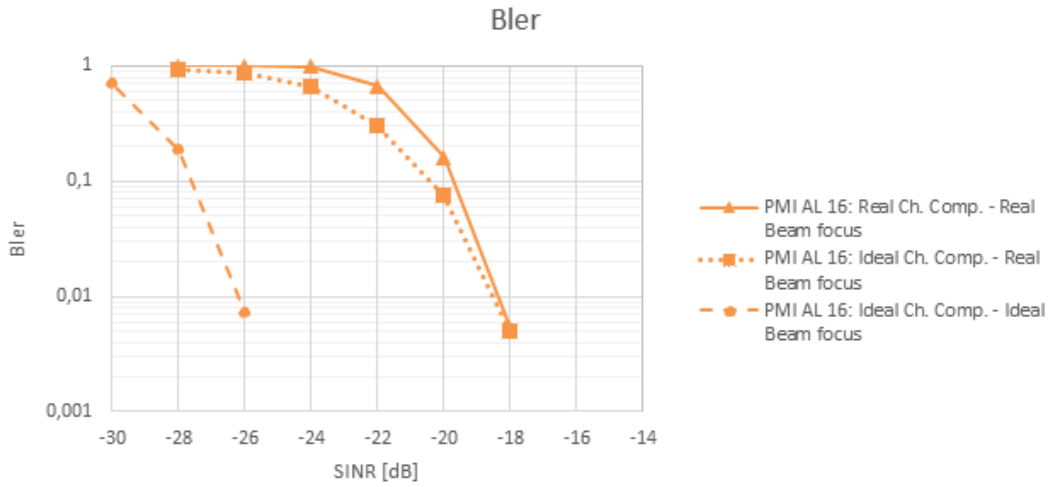


Figure 5.9 *Follow PMI: comparison between ideal and real channel compensation*

From the obtained results, it can be deduced that the performances of the follow PMI approach can be enhanced by improving the pointing of the beam and it is possible with a better estimate of the channel obtained from the CSI-RS.

Eigenvector beamforming: ideal vs real behavior

The behavior of eigenvector beamforming with non-ideal channel estimation is reported in Figure 5.10.

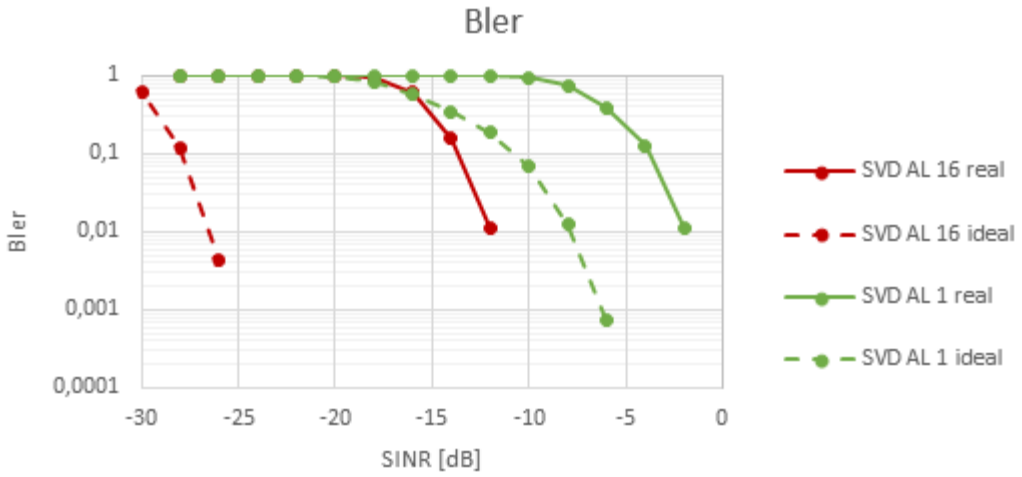


Figure 5.10 *Eigenvector beamforming performances in ideal and real scenarios*

With the eigenvector beamforming the differences between real and ideal cases are even more pronounced since the beamforming is designed on the basis of the channel matrix H that is estimated through the SRS transmitted by the UE which has a limited amount of available power.

In fact, while in the follow PMI the channel estimate is built using CSI Reference Signal that are sent by the gNB, in the eigenvector beamforming the channel estimate is built using Sounding Reference Signal sent by the UE. However, the gNB operates with a much higher power (typically $46\ dBm$ every 20 MHz) compared to the one available in battery limited UEs (typically only $23\ dBm$)

This limitation in the power available for the reference signals leads to a worse channel estimation and consequently a worse pointing of the beam.

Results in Figure 5.7 verify an aspect detected also for transmissions over the PDSCH and that is related to the usage cases of the two approaches. In fact, it can be clearly noticed that performances obtained in Figure 5.7 corresponds to low values of SNR, and for both aggregation levels 1 and 16 the follow PMI performs better than the eigenvector method. This leads to the conclusion that in conditions where SNR values are low it is better to adopt the follow PMI approach, as in the case of terminals in proximity of cell edge.

Instead, for higher SNR it is preferable to use the eigenvector beamforming.

6. Conclusions

This document is aimed to be a report of the project carried out in Telecom Italia about the topic of the *Physical Downlink Control Channel* of 5G.

In the initial part of the project, it has been necessary to learn the theoretical part of the physical layer for the mobile communication systems of the fifth generation. Standards published by 3GPP have been the main source for this purpose.

After this preliminary study, the implementation of the simulator software has followed. The simulator is developed through Matlab tools and reproduces the block chain for the processing of control information over the PDCCH. The most relevant modules in the channel model are the module for the polar encoding/decoding, the module for the MIMO processing, for the channel estimation and compensation.

The obtained results have demonstrated that in real applications, where the channel estimation is based on reference signals actually exchanged between gNB and UE, the design of beamforming is strongly compromised when the Signal to Interference ratio is very low. As a consequence, it is fundamental the employment of techniques for the channel estimation as more accurate as possible. Another feature that is affected by the real channel estimation is the compensation of the channel, but results show that this effect is less important compared to the error due to a compromised beamforming.

Simulations for different values of the “aggregation level” parameter have illustrated the benefit of a larger physical resource allocation: when the PDCCH is assigned a larger amount of resources, a higher number of parity bits are available, hence the control information is more protected.

Further observations have allowed to compare two possible approaches for the beamforming. The first one is based on reference signals sent by the gNB (the CSI-RS), while the second one is based on reference signals transmitted by the UE (SRS). In the first case, the beamforming has been designed through a method called *follow PMI* while, for the second, the eigen-beamforming approach has been applied. It has been noticed that the follow PMI approach performs better than the eigenvector beamforming, due to the limited available power at the terminal side.

The results obtained provide SNR working point for the PDCCH control channel, that is a fundamental input to the planning and link budget tools employed for the management of the mobile networks in Telecom Italia.

7. Bibliography

[1] TELCOMA, 5G Technology Introduction, 2018.

[2] R12-SG05, “IMT Vision - "Framework and overall objectives of the future development of IMT for 2020 and beyond",” 2015. [Online]. Available: https://www.itu.int/rec/R-REC-M.2083-0-201509-I/en.

[3] E. Dahlman, S. Parkvall and J. Skold, 5G NR: The Next Generation Wireless Access Technology, Academic Press, 2018.

[4] 3GPP, “4. Services and functions of the physical layer,” in *3GPP TS 38.202 V15.2.0 (2018-06)*.

[5] 3GPP, “Part 1, Part 2,” in *TS 38.101: NR; User Equipment (UE) radio transmission and reception*, 2019-02-14.

[6] 3GPP, “4.2 Numerologies,” in *TS 38.211 Physical channels and modulation V15.2.0*, 2018-06.

[7] 3GPP, “4.3 Frame structure,” in *TS 38.211 Physical channels and modulation V15.2.0*, 2018-06.

[8] 3GPP, “4.4 Physical resources,” in *TS 38.211 Physical channels and modulation V15.2.0*, 2018-06.

[9] 3GPP, “5 Generic functions,” in *TS 38.211 Physical channels and modulation V15.2.0*, 2018-06.

[10] “Wikipedia,” [Online]. Available: https://en.wikipedia.org/wiki/Orthogonal_frequency-division_multiplexing.

[11] 3GPP, “7 Downlink, 6 Uplink,” in *TS 38.211 Physical channels and modulation V15.2.0*, 2018-06.

[12] 3GPP, “7.3.2 Physical downlink control channel (PDCCH),” in *TS 38.211 Physical channels and modulation V15.2.0*, 2018-06.

[13] 3GPP, "7.3 Downlink control information," in *TS 38.212 Multiplexing and channel coding (Release 15) V15.2.0*, 2018-06.

[14] 3GPP, "5 General procedures," in *TS 38.212 Multiplexing and channel coding (Release 15) V15.2.0*, 2018-06.

[15] 3GPP, "7 Downlink," in *TS 38.211 Physical channels and modulation V15.2.0*, 2018-06 .

[16] Erdal Arıkan, Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, 2009.

[17] K. Niu, K. Chen, J. Lin and Q. T. Zhang, Polar Codes: Primary Concepts and Practical Decoding Algorithms, IEEE Communications Magazine, July 2014.

[18] H. Vangala, Y. Hong, E. Viterbo and A. Monash University, "A Practical Introduction to Polar Codes," 23 February 2016. [Online]. Available: is.gd/polarcodes.

[19] I. Tal and A. Vardy, List-Decoding of Polar Codes, 9500 Gilman Drive, La Jolla, CA 92093, USA: University of California San Diego.

[20] I. Tal and A. Vardy, List Decoding of Polar Codes, IEEE, 2015.

[21] 3GPP, TS 38.214 V15.2.0 Physical layer procedures for data (Release 15), 2018-06.

Acknowledgments

At the end of this document, I feel it my duty to acknowledge all the people who have stood by me throughout my years at university and not only.

My sincere gratitude is to my parents, Angelo and Lina, for imparting me morals and the worth of culture, for giving me the opportunity to study and for supporting me in all the ways.

Thanks to my sister, Viviana, for always believing in me more than I do.

I also wish to thank all my relatives who have supported me and rejoiced in my accomplishments.

I'm grateful to everyone, both near and far, who has contributed to the achievement of this goal, in particular Lillo, Edera, Claudia, Carla, Margareth, Cynthia, Gera and Spin.

Thanks to my university colleagues for all good but also bad times spent together in preparing exams, projects and laboratories.

I'd like to thank my supervisor, prof. Roberto Garello, model of erudition and humanity, which has always been available and paid attention to my needs.

Finally, but not least, I'd like to express my warm thanks to my tutor in Telecom Italia, Roberto Fantini, without which this thesis project would not have been possible. I'm thankful to him for having guided me since the beginning sharing with me, patiently, his huge knowledge, for always being available to help me, providing precious hints.